

INFERENCE, THOUGHT EXPERIMENTS, AND PHYSICALISM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Stephen Mahaffey May 2014

© 2014 Stephen Mahaffey

# INFERENCE, THOUGHT EXPERIMENTS, AND PHYSICALISM

Stephen Mahaffey, Ph. D.

Cornell University 2014

This dissertation addresses and connects three issues. First, I seek to better understand philosophers' use of thought experiments. In particular, I seek to better understand the origin and epistemic status of the conclusions we glean from such experiments. To do that, I address the questions: How do we even form these conclusions? And are they generally reliable and well justified? If so, why and how? In response, I develop a new theory of how we form our conclusions. The core idea is that we infer them from the suppositions we make to create the imaginary cases in our thought experiments. This theory entails that our conclusions about thought experiments are reliable and typically bear a priori justification.

Another key issue is this: if we know all of the fundamental truths of reality, and thus know not only the fundamental laws of nature but also the location, interrelations, and intrinsic properties of every fundamental entity, can we (assuming we're ideally smart and energetic) then infer every other truth of reality? The nub of this issue is that the fundamental truth metaphysically explains all other truths. So what's really at stake is whether metaphysical explanations are *transparent*: viz., open for us to see how *explananda* follow from *explanans*, and so determine the former from the latter.

I argue that we cannot infer all non-fundamental truths from the fundamental truth—and, so, that metaphysical explanations sometimes are not transparent. To argue for this conclusion, I work with the standard assumption that the fundamental truth concerns microphysical reality. I

then argue that there are many non-fundamental macrophysical truths that even idealized agents cannot infer from the fundamental microphysical truth.

The final issue I address is the nature and origin of phenomenal consciousness—i.e., the subjective character of our experiences that we try to describe with statements about what our experiences are like for us. On this topic, I defend *physicalism* about phenomenal consciousness against several recent and influential objections. This means, for one thing, that I defend the view that a person's conscious experience at any given time is determined by her neurobiological state at that moment.

## BIOGRAPHICAL SKETCH

Stephen Mahaffey began to study philosophy at the University of Calgary, where he earned a B.A in philosophy in 2007. In 2014, after seven wonderful years at Cornell, he earned a Ph.D in philosophy with this very document.

To my Mother

## ACKNOWLEDGMENTS

First, I want to thank the member of my committee: Ted Sider, Nico Silins, Karen Bennett, and, especially, Derk Pereboom. I am greatly indebted to them for their conversation, comments, advice, warmth, and patience. I would not have been able to write this dissertation without their help. I want to particularly thank Derk for his unfailing guidance, wisdom, and encouragement. My debt to him goes far beyond the writing of this dissertation.

I also want to thank a number of friends at Cornell, who read my dissertation as it progressed and offered helpful feedback. Thus, thanks to Theodore Korzukhin, Scott O'Connor, Colin McLear, Subrena Smith, and Clifford Roberts. Special thanks go to Adam Bendorf and Jonathan Vance, who have helped me enormously with my work, and have been both constant allies and the best of friends. You guys are wonderful.

I want to thank the Cornell philosophy community for being such a friendly and supportive group of people. Thank you for listening to my talks, and providing such helpful feedback.

Finally, I want to thank my Mother for her tremendous love, care, joy, and support. You're a treasure Mom.

## TABLE OF CONTENTS

Biographical Sketch_____	iii
Dedication_____	iv
Acknowledgements_____	v
Table of Contents_____	vi
 <i>Introduction</i> _____	 1
 <i>Chapter 1: A Priori Abduction and Philosophical Thought Experiments</i> _____	 5
1.1 Abduction and Birds of Paradise_____	10
1.2 Hypothetical Reasoning and Conditional Beliefs_____	12
2.0 A Priori Justification_____	14
2.1 Concept Possession_____	16
2.2 Justication for BP_____	16
2.3 What About Background Beliefs?_____	17
2.4 Do We Rely on our Past Experience with Successful Abduction?_____	18
2.5 Fallibility_____	21
3.0 Thought Experiments_____	22
3.1 The Plan_____	23
3.2 The Theory: In Brief_____	24
3.3 More Details: Imagining the Case_____	25
3.4 More Details: Drawing the Inference_____	28
3.5 The Contents of our Judgments_____	31
4.0 A Priori Justification_____	33
4.1 Don't we Rely on A Posteriori Knowledge about the Actual World?_____	33
4.2 What about Background Beliefs?_____	34
4.3 Do we Rely on Evidence that Abduction is Reliable?_____	36
5.0 A General Theory_____	37
5.1 Imaginary and Non-Imaginary Experiments_____	39
5.2 Reliability and Jusitification_____	39
6.1 Synthetic A Priori Knowledge_____	40
6.2 Philosophy <i>Sans</i> Intuitions_____	42
 <i>Chapter 2: Inference and the Epistemology of Grounding</i> _____	 43
1.1 Importance of Inferentialism: Metaphysics_____	48
1.2 Other Implications_____	48



2.0 The Precedent for Inferentialism	50
3.1 The Case for Inferentialism: A Positive Argument	52
3.2 Scarce Counterexamples	52
3.3 Controversial Counterexamples	53
4.1 The Case Against Inferentialism: Macrophysical Truths	54
4.2 Extant Arguments and Setup	56
4.3 Interpretation #1	57
4.4 Are We Done Already?	61
4.5 Problems with Extant Arguments	64
5.0 Interpretation #2	66
5.1 Many Theories of Concept Possession	67
5.2 Undemanding Theories	68
5.3 An Important Qualification	71
5.4 A More Demanding Theory: Recognitional Concept Possession	72
5.5 A More Demanding Theory: Inferential Concept Possession	73
5.6 Conclusion: Concept Possession Does not Help	76
5.7 Bringing it all Together	77
6.1 An Alternative Thesis?: First Try	78
6.2 Second Try	80
<i>Chapter 3: Mary Macrophysical Ignorance</i>	84
1.0 The Progression of the Knowledge Argument: Two Formulations	86
2.0 Advancing the Dialectic	89
2.1 Old Facts	89
2.2 New Abilities	91
3.0 ~InferQ and InferM	93
4.0 An Overlooked Opportunity	94
4.1 Incomplete Physical Knowledge Response	95
4.2 The Indexical Reply	95
4.3 Acquaintance Reply	96
4.4 Missing-Concept Reply	97
5.0 A New Objection	98
5.1 P Alone	98
5.2 Enabling Information	99
5.3 Changing Cognitive Roles	101
5.4 Mastering M-Concepts	103
5.5 Mastery as Recognitional Skill	104

5.6 Mastery as Inferential Skill	105
5.7 Mastery as Knowledge	107
6.0 Adding to Mary's Justifying Information	109
6.1 A Futile Change	110
6.2 Conclusion	111
<i>Chapter 4: Zombie Worlds and Strong Necessities</i>	112
1.1 Clarifying the Discussion: Zombies?	114
1.2 The Crucial Second Premise	116
2.0 A Complex Dialectic	117
2.1 A Priority	117
2.2 Ideality	118
2.3 Imagination	120
2.4 Dimensions of Possibility	123
2.5 Two-Dimensional Semantics	124
2.6 Two-Dimensional Modality	126
2.7 A Two-Dimensional Argument	127
3.0 The State of the Opposition: Strong Necessities	128
4.0 PTI $\supset$ M Conditionals	130
4.1 Primarily Necessary	130
4.2 Not A Priori	130
4.3 Imaginable as False	131
4.4 Conclusion	135
5.0 A Response from Modal Theory	136
5.1 Modal Monism	136
Bibliography	142

## **Introduction**

This dissertation addresses and connects a number of questions. One group concerns the nature and origin of phenomenal consciousness—i.e., the subjective character of our experiences that we try to describe with statements about what our experiences are like for us. This focus presents questions like: is phenomenal consciousness a phenomenon we can come to fully understand through physical science? And is the occurrence and character each person's individual consciousness explained by her neurobiological state and her relations to the physical world?

Another key question is this: if we know all of the fundamental truths of reality, and thus know not only the fundamental laws of nature but also the location, interrelations, and intrinsic properties of every fundamental entity, can we (or at least idealized versions of us) then infer every other truth of reality?

The final set of questions concerns philosophical methodology; in particular, our use of thought experiments—such as Gettier cases, Frankfurt cases, Twin-earth cases, etc. My primary focus here is the origin and epistemic status of the judgments we make about such experiments—such as our judgment about how the protagonist in a Gettier case truly, justifiably believes but does not know. Relevant questions include: How do we even form these judgments? Are these judgments generally reliable? Are they typically well justified? If so, why are they reliable, and how do they acquire justification? And what sort of justification do they typically acquire?

I argue for answers to all of these questions, and show how they're deeply interconnected.

I start with philosophical methodology in chapter 1, where I defend the reliability and justification of our judgments about thought experiments, and, in doing so, try to reconcile two *prima facie* plausible positions that present consensus takes to be incompatible. These are i) the view that our judgments are *a priori* justified when justified at all, and ii) skepticism about the existence and justificatory power of intuitions. To do this, I develop a new account of how we form and justify our judgments: one that makes them *a priori* and yet does not posit any role for intuitions. The core idea of this account is that we infer our judgments from the suppositions that we make to imagine the cases in our thought experiments.

In chapter 2 I argue that we cannot infer all non-fundamental truths from the fundamental truth. That is, I argue that even if we know all of the fundamental truths of reality, and thus know not only the fundamental laws of nature but also the location, interrelations, and intrinsic properties of every fundamental entity, there are still some non-fundamental truths that we cannot thence infer. This means that even though the fundamental truth metaphysically explains all non-fundamental truths, this explanation often fails to be *transparent*—viz., is often such that we cannot determine explanandum from explanans. To argue for this conclusion, I work with the standard assumption that the fundamental truth concerns microphysical entities, their properties and interrelations, and the laws that govern them. I then argue that there are many non-fundamental macrophysical truths that even idealized agents cannot infer from the fundamental microphysical truth.

My final topic is phenomenal consciousness. Here I draw on my previous arguments to defend *physicalism*—viz., the view that phenomenal consciousness is something we can come to understand through physical science, and that each person's individual consciousness arises from

her neurobiological state and her relations to the physical world. In particular, I develop new responses to two of the most influential and powerful arguments against physicalism: the i) knowledge argument, and ii) conceivability argument.

Chapter 3 discusses the knowledge argument. This argument starts with a famous thought experiment. Our protagonist is Mary and she's never experienced colour because she's lived her whole life in a black and white room. However, she not only has access to every colourless book, lecture, and TV program in the world, but, through her phone, can converse with any scientific researcher in the world. Nonetheless, the knowledge argument claims that it's impossible for Mary to learn phenomenal truths about what it's like to see colour unless she escapes her room. This case is supposed to illustrate that it's possible to have epistemic access to all of the physical truths of science while having no access to certain phenomenal truths (the argument usually focuses on truths about what it's like to see colour). On this basis, the knowledge argument contends that phenomenal truths are not physical truths. In short, that phenomenal truths and the consciousness they describe are beyond the ken of physical science.

In response, I argue that Mary's ignorance of phenomenal truths about what it's like to see color is matched by her ignorance of many perfectly ordinary macrophysical truths. This means that if we set Mary's case such that she cannot learn certain phenomenal truths while trapped in her black-and-white room, then she likewise cannot learn many macrophysical truths. And if we ensure that she can learn all macrophysical truths, then we also equip her to learn relevant phenomenal truths. The end result is that either Mary knows what it's like to see colour, or her ignorance of phenomenal truths does not show that such truths—and the properties they concern—are a class apart from physical truths and properties. Either way, the knowledge argument is in

trouble.

Finally, chapter 4 discusses the conceivability argument. This argument works by trying to show that there are possible worlds that are microphysical duplicates of the actual world, and yet different at the level of phenomenal consciousness—these worlds are often called zombie-worlds. For if such worlds are possible, it follows that microphysical truths and properties do not in fact explain the occurrence and character of phenomenal truths and properties. To get the desired modal conclusion, the argument contends that since zombie-worlds are conceivable, it follows that they're possible.

In response, I argue that we can and should resist the idea that conceivability entails possibility when it comes to zombie worlds. To do this I argue that we can likewise conceive of worlds that are microphysical but not macrophysical duplicates of our world. I then argue that such 'macro-zombie' worlds are altogether impossible. Together these two conclusions show that, in general, conceivability does not entail possibility. This result blocks the claim that the possibility of a zombie world follows from its conceivability.

My discussion of the conceivability and knowledge arguments focuses on recent iterations that incorporate two-dimensional modal and semantic theory, and therefore employ the two-dimensionalist distinction between (at least) two varieties of possibility, necessity, and representational content. I intend my responses to be effective against even these new and sophisticated versions of the arguments in question. So, for instance, I argue that even if zombie worlds are conceivable, this does not show that they're possible in any of the ways posited by two-dimensional theory.

## Chapter 1: A Priori Abduction and Philosophical Thought Experiments

It almost goes without saying that we philosophers often rely on thought experiments to drive our arguments.<sup>1</sup> For it's standard practice to test philosophical hypotheses and theories against imaginary cases. We do this by dreaming up a case, carefully forming an intuitive judgment about it, and then checking to see whether our judgment fits with the theory we're testing.<sup>2</sup> If not, we usually take this mismatch to be evidence against the theory, and sometimes we even reject it on these grounds. To see this methodology at work, consider the way we use imaginary Gettier cases. When we consider such test-cases, we judge something like “In this case, Smith truly, justifiably believes but does not know”, and we take this judgment to be conclusive evidence against the theory that knowledge is true, justified belief.

While use of thought experiments is standard practice in philosophy, there is much we do not understand about this practice. We're still trying to answer questions like: How do we even form our judgments about thought experiments? Are they generally reliable? Are they typically justified? If so, why are they reliable, and how do we acquire justification for them? These questions matter: if our judgments are typically unjustified and/or unreliable, or if we cannot satisfactorily explain how we might form justified, reliable judgments, then we should be suspicious about our practice of consulting thought experiments.

In response to such questions, in this paper I develop a new theory of how we form and justify our judgements about thought experiments. To introduce this theory and highlight its

---

<sup>1</sup> And philosophers aren't the only ones. Thought experiments are widely employed in disciplines like physics and economics, and sometimes even in history (I'm thinking of the controversial practice of counterfactual history here). In this paper, I focus on thought experiments in philosophy, but I suspect that everything I say also applies to thought experiments from other disciplines. I leave this issue to future work.

<sup>2</sup> Some people use the term “intuitions” where I use “judgements”, but as I'll soon explain, I want to reserve the former term for another use.

novelty, let's first locate it in the extant dialectic. To that end, there are four key debates that dominate present discussion and helpfully organize the dialectic. The first is the debate on whether our judgments about thought experiments are generally reliable and typically justified. The second concerns whether the cognitive processes that produce our judgments are of a kind with the processes at work when we form judgements about non-imaginary experiments and cases. The third is about whether our judgments are, when justified at all, a priori or a posteriori justified. Finally, the fourth debate concerns whether our judgments about thought experiments are occasioned and justified by our *intuitions*: viz., by *sui generis* conscious, contentful intellectual seemings that i) are not judgments, introspections, or perceptions, ii) have a distinctive phenomenology, and iii) arise purely from our understanding of the proposition intuited, not from introspection, perception, memory, or inference.

The dialectic coalesces into three main camps around the above debates. First, there are (and here I use my own unofficial monikers) the *skeptics*: who—often on the basis of results from experimental philosophy—cast doubt on the reliability and justification of our judgments about thought experiments (see, e.g. Hintikka 1999; Weinberg, Nichols, & Stich 2001; Kornblith 2002, 2005, 2006; Weinberg 2007; Weinberg, Gonnerman, Buckner, Alexander 2010; Buckwalter & Stich 2011).<sup>3</sup> Next, there are the *intuitionists*: who argue i) that our judgments are generally reliable and justified, ii) that they're typically a priori justified when justified at all, and iii) that we do rely on our intuitions to justify and prompt our judgments (see, e.g., Bonjour 1998; Pust 2000; Bealer 2000, 2002; Sosa 2007a, 2007b, 2009; Chudnoff 2012). Third, there are the *naturalists*: who argue i) that our judgments are generally reliable and justified, ii) that

---

<sup>3</sup> Unsurprisingly, these writers do not have much to say about the other two debates listed above.



our judgments are not a priori justified, and iii) that we do *not* rely on intuitions as they're construed above (see, e.g., Goldman 2007; Williamson 2008; Nagel 2012; Boyd 2013).<sup>4</sup> Naturalists typically hold that our judgments about thought experiments arise from the same cognitive processes that are at work when we assess and form judgments about non-imaginary experiments and cases.

One major theme now stands out: the people who presently argue that our judgments about thought experiments are typically a priori justified (when justified at all) also argue that our intuitions occasion and justify our judgments. In fact, this correlation is so universal that it's commonly assumed that if one upholds the a priority of our judgments, one also defends a justificatory role for intuitions. So according to the present dialectic, the view that our judgments are a priori justified is incompatible with an approach that does not posit such a role for intuitions.

This is a suboptimal state of affairs. For both of these supposedly incompatible positions are *prima facie* plausible. On the one hand, it's *prima facie* plausible that our judgments about thought experiments are a priori because such experiments feature imaginary cases; cases that we evaluate, it seems, by simply thinking about their imagined nature. Since we evaluate such cases in thought, it's natural to think that any justification we obtain for our subsequent judgments derives from rational reflection, and is thus a priori. What's more, it's hard to see how empirical evidence or grounds might help us to make one judgment rather than another about a thought experiment. For instance, what can empirical evidence tell us about the protagonist in a wholly

---

<sup>4</sup> Though naturalists often present themselves as defending our intuitions (see, for instance, Goldman 2007), what they mean by this is that they're defending the reliability of our judgments about thought experiments. Unhappily, "intuition" is used in multiple ways: sometimes to refer to the seemings I (and many others) call intuitions, and sometimes to the judgments that may or may not be justified by intellectual seemings.

*imaginary* Gettier case?

On the other hand, it's reasonable to be suspicious of the idea that we have intuitions which prompt and justify our judgments about thought experiments. For as many writers have pointed out (e.g., Field 1989; Boghossian 2003; Wright 2004b; Malmgren 2011), no one has yet been able to explain exactly how we form intuitions, why they're reliable, and why they have justificatory power. In short, the suggestion that we rely upon our justificatory intuitions is—at least in the mind of many philosophers—rather mysterious. What's more, other philosophers further contend that there's no reason to think that intuitions exist at all (Williamson 2008, Cappelen 2012). Proponents of intuitions of course try to dispel these worries—and/or convince us that they don't matter—but, at the very least, intuitions are a controversial and uncertain basis for an epistemology of thought experiments. We therefore have good reason to look for and, if we're not skeptics, esteem alternative theories that do not posit a role for intuitions.

In this paper I argue that in fact we can accept both of the *prima facie* plausible views I just presented. That is, I try to show that, *pace* present consensus, the view that our judgments about thought experiments are a priori justified (when justified at all) is compatible with an approach that rejects a justificatory role for intuitions. This is because, I argue, we do not need to invoke intuitions to give a compelling account of how we might a priori justify our judgments. Alternatives are available. To make this evident, I propose and develop a new account, one which entails that our judgments are typically a priori justified when justified at all, and yet dispenses with any role for intuitions. This theory also allows us to hold that even though our judgments are typically a priori justified, they're the product of a cognitive process that's also at work when we draw conclusions about the non-imaginary experiments we conduct in scientific

inquiry and everyday life. The result is a theory that combines key ideas from what I've called intuitionism and naturalism, and thereby carves out a new place in the dialectic; one we might describe as *a priori naturalism*.

The core idea of this theory is that we can *infer* our judgments about thought experiments, and that our premises are just the suppositions we make when we conceive of the imaginary cases at hand. That is, I propose that when we consider the test case in a thought experiment, we can reach a judgment by inferring on the basis of the suppositions we make to craft this experiment. It's this inference that, when rational and unchallenged by defeaters, yields a priori justification for our resulting judgment. What's more, I argue that the inference pattern in question is inference to the best explanation (henceforth, abduction). Thus, on the theory I develop, we can infer our judgments via a process of abduction that, when all goes well, yields a priori justification for our resulting judgments. In short, I submit that we can *a priori abduce* our judgments about thought experiments.

If the abductive theory I suggest is compelling, then it presents us with the a priori naturalist position that I seek. This is because, as I'll argue, i) abduction is a generally reliable pattern of inference, ii) rational inferences on the basis of suppositions are, absent defeaters, a good source of a priori justification, and iii) if our capacity for abduction can explain how we (a priori) justify our judgments about thought experiments, then we can dispense with the idea that intuitions are what prompt and justify such judgments. What's more, this theory harmonizes the processes we use to investigate thought experiments with the processes we use to investigate non-imaginary, empirical experiments. For it's clear that we often use abduction to draw conclusions about the non-imaginary experiments we conduct in science and everyday life.

This paper splits into two parts. The first half presents an independent argument for the conclusion that some of our abductions yield a priori justification for our resulting conclusions. That is, I independently argue that a priori abduction is an actual phenomenon. This argument prepares the ground for the theory to come.<sup>5</sup>

The second half of the paper builds on my initial argument to develop an account of how we a priori abduce our judgments about thought experiments. To make matters less abstract, I focus on explaining one way that we might a priori abduce our 'no-knowledge' judgement about a Gettier case. I also argue that the final prospect of this paper is a novel picture of philosophical methodology: one that dispenses with intuitions as sources of justification, but retains the idea that philosophical reasoning often yields a priori justification and synthetic a priori knowledge.

## 1) Abduction and Birds of Paradise

Consider the following set of claims about sexual selection and King of Saxony Birds of Paradise.<sup>6</sup> Call this set, **Birds of Paradise (BP)**:

---

<sup>5</sup> This argument is also my attempt to dispel what, at present, is widespread scepticism about the actuality of a priori abduction. In fact, the only people who come presently close to my position are Hawthorne (2002), Cohen (2010), and Wedgwood (2012). Hawthorne suggests that a priori abduction is possible, but, crucially, he doesn't argue that it's actual. Cohen and Wedgwood argue for the actuality of what Wedgwood calls "a priori bootstrapping". This is a "non-deductive" process of inference that yields a priori justification for the material conditional *If I cannot detect that my sensory experiences are unreliable  $\supset$  my sensory experiences are reliable*. This conclusion sounds very similar to mine, but, interestingly, neither Cohen nor Wedgwood rule on whether their non-deductive inference is abductive, so it may be that we do not agree after all.

<sup>6</sup> As I mention in note 5, I know that Hawthorne (2002) also presents an example of what might be a priori abduction. However, his example does not suffice for my purposes. This is because even if his example works (*pace* critics like Turri (2011)) it can only show that a priori abduction is possible. It cannot show that a priori abduction is actual—which is what I need to establish in order to argue that we a priori abduce our judgments about thought experiments. So in order to further my aims, I need to devise a new example that's importantly different from Hawthorne's.

Why? Well, Hawthorne's example concerns The Explainer. The Explainer is a yet-to-be-embodied being who, prior to having any sense experiences, thinks about which microphysical theories best explain the potential experiential life-histories that she conceives of for herself. Hawthorne argues that The Explainer can gain a priori justification for beliefs about how some microphysical theory, T, best explains some potential life-history, L. Hawthorne then suggests that since The Explainer has a priori justification to believe that T best explains L,

1) Male King of Saxony Birds of Paradise (whose bodies range from 20-25 cm long) sport magnificent brow plumes that grow up to 50 cm long.

2) The best explanation of (1) is that male Kings of Saxony evolved such plumes through a process of sexual selection.<sup>7</sup>

Now consider:

**Sexual Selection (SS):** Male King of Saxony Birds of Paradise evolved magnificent, 50cm long brow plumes through a process of sexual selection.

It's clear that we—i.e., actual people—can rationally infer SS from BP. In fact, this is the inference that ornithologists drew when they reached the now established conclusion that male King of Saxony Birds of Paradise (like all male birds of paradise) owe their fantastic plumage to the evolutionary process of sexual selection.

But when we infer SS from BP, our inference is not deductive. BP does not logically entail SS. Instead, this inference is a clear case of abduction. For here we infer SS because, given BP, sexual selection best explains the brow plumes of male King of Saxony Birds of Paradise. This is much the sort of inference to the best explanation that we perform all the time in both our daily lives and our scientific enquiry. The only unusual aspect of my example is that BP explicitly states that sexual selection is the best explanation of the relevant brow plumes. So when we suppose that BP is true, we don't have to figure out the best explanation of such plumes in order to then infer it. We can instead immediately abduce the conclusion—viz., SS—that BP

---

she can justifiably abduce the material conditional *If I have L  $\supset$  T is true*. Hawthorne's idea is that since The Explainer's abduction proceeds from a priori justified premises—and is performed by a being who's yet to have any sense experiences—it provides her with a priori justification for *If I have L  $\supset$  T is true*. The overall result: it seems as if The Explainer performs an a priori abduction.

From my perspective, the problem with Hawthorne's example is that actual people cannot perform The Explainer's fantastic inference, do not reflect prior to having sense experiences, and, arguably, cannot even grasp a complete microphysical theory or a full experiential life-history.

<sup>7</sup> Sexual selection is the mode of natural selection where a population evolves a trait because it makes exhibitors more attractive to the opposite sex, and ensures that individuals with this trait reproduce far more than those who lack it.

anoints as the best explanation.

## 1.2 Hypothetical Reasoning and Conditional Beliefs

When ornithologists abducted SS from BP, they did so with knowledge of BP. But it's important to see that we do not need such knowledge in order to rationally draw this inference. We do not even need to believe BP. This is because we can *hypothetically* abduce SS from BP: i.e., we can suppose that BP is true, and then rationally abduce SS on the basis of our supposition. In this paper, I will focus on our capacity to draw this hypothetical inference from BP to SS. For it's this hypothetical abduction that's a good candidate to be an a priori abduction.

When we hypothetically abduce SS from BP, our inference of course does not provide us with justification to believe SS. For in order to get such justification, we'd first have to gain justification to believe BP—just like ornithologists did. Nonetheless, I argue that our hypothetical abduction does yield some justification. In particular, it yields justification, absent defeaters, to believe *If BP, then SS* and *If BP were true, then SS would be true*—henceforth, *BP > SS*. This is because, as our everyday practice of hypothetical reasoning makes clear, when we make a supposition and then rationally infer a conclusion on this basis, we typically gain justification to discharge our supposition and believe that if it's (were) true, then so is (would be) our conclusion.<sup>8</sup> This is so regardless of whether our inference is deductive, abductive, inductive, or otherwise.

Many examples from everyday reasoning support this position. Let's first look at examples where hypothetical reasoning provides people with justification for conditional beliefs.

---

<sup>8</sup> In everyday reasoning, the matter of whether we form a conditional or a counterfactual belief is determined by whether we regard our supposition to be counterfactual or not.

Suppose that I'm friends with Sergei, who was once a longtime park ranger on Wrangel Island, serving there for fifteen years. Now also suppose that someone else asks me whether it will be buggy on Wrangel this upcoming August. In response, it's clear that I can justifiably reply,

“Ask Sergei. If it was buggy every August that Sergei was on Wrangel, then it will be buggy this August”.

I arrive at my conditional belief here through a process of hypothetical induction: I suppose that all of Sergei's Augusts were buggy, and then on that basis induce, rationally it seems, that this August will likely be buggy too. I then discharge my supposition and judge, with justification, that if it was buggy every August Sergei was on Wrangel, then it will be buggy this upcoming August.

As another example, here's a true story. Recently I heard footsteps, rustling, and gnawing noises in the attic above my apartment. Alarmed, I called my landlord to report the matter. I asked him, “What type of animal is living in the attic?” and “How did they get in?” He answered,

“Well, if it's squirrels, they got in from outside”  
and

“Whereas, if it's rats, they got in from the sewer, or from a neighbour's place”

Given his wealth of experience, it's plausible that my landlord's beliefs are justified. But how then did he reach them ? It seems clear that he used hypothetical abduction. Take his squirrel belief: in this case my landlord supposed that squirrels are in my attic, drew on background knowledge about their habits, and then rationally abduced that the squirrels got in from outside. He then discharged his supposition and judged, with justification, that if it's squirrels in my attic, they broke in from outside. The story is essentially the same for my landlord's rat belief. Given

such examples, it's plausible that when we suppose BP and then rationally abduce SS on the basis of this supposition, we can similarly discharge our supposition and, absent defeaters, justifiably believe *If BP, then SS*.

Now let's look at an example where non-deductive hypothetical reasoning yields justification for a counterfactual belief. Imagine a professor of geology teaching an introductory class on plate tectonics. Now suppose that a student asks our geologist what the world would be like if the Hawaiian hotspot had never existed. Our geologist will reply,

“If the Hawaiian hotspot had never existed, then the Hawaiian islands would not exist”

Given our protagonist's expertise—and the nature of the earth—this belief is surely justified. But how then does he reach it? He does this via hypothetical abduction. He forms the counterfactual supposition that the Hawaiian hotspot does not exist, draws on his expertise for further suppositions to fill in the background of his imaginary counterfactual scenario, and then, on the basis of these suppositions, rationally abduces that the Hawaiian islands would not exist. He then discharges his suppositions and arrives at his well-justified counterfactual belief from above. This commonplace example strongly suggests that we can reason in much the same way with BP and SS. That is, it's plausible that when I suppose BP and thence abduce SS, I, like our geologist, can then discharge my supposition and, absent defeaters, justifiably believe  $BP > SS$ .

## **2) A Priori Justification**

I've now argued that when we hypothetically abduce SS from BP, we gain justification to believe *If BP, then SS* and  $BP > SS$ . But what sort of justification do we acquire? Is it a priori or a posteriori? I argue that this justification is a priori.



I will work with a simple and intuitive characterization of a priority: justification is *a priori* just in case it derives solely from rational reflection and, so, does not even partly derive from sense experience or introspection. As such, to argue that our justification to believe *If BP, then SS* and  $BP > SS$  is *a priori*, I argue that it derives solely from rational reflection, as sense experience and introspection do not play any sort of justificatory role for us when we hypothetically abduce SS from BP.

I'll start with a quick case for the conclusion that when we hypothetically abduce SS from BP, the resulting justification to believe *If BP, then SS* and  $BP > SS$  derives solely from rational reflection. Let's think about what's involved with the inference of SS from BP. One way to do this is to parse the various components of this inference. This gives us a list like: i) suppose that BP is true, ii) understand BP, iii) understand SS, iv) realize that, given BP, sexual selection best explains the plumes of male King of Saxony birds of paradise, and, finally, v) hypothetically abduce SS.<sup>9</sup> But notice, all of these component processes look like processes of rational reflection. Supposing, understanding, realizing, inferring; these are all modes of thought. So it appears that we infer SS from BP by engaging in a multifaceted process of rational reflection. And since this inference is what supplies us with justification to believe *If BP, then SS* and  $BP > SS$ , we therefore have at least *prima facie* reason to think that our justification derives solely from rational reflection. That is, we've *prima facie* reason to think that our justification is *a priori*.

I think the above argument is compelling. But, to press this case, I'll now head off various reasons for thinking that the justification provided by our inference of SS is *a posteriori*.

---

<sup>9</sup> In numbering these steps, I do not intend to put them in chronological order—for at least some of these steps could occur simultaneously. I merely intend to label the various components of this inference.

In doing so, I'll argue that a posteriori justification does not play any role in our abduction of SS.

## 2.1 Concept Possession

One reason is that we need to have appropriate experiences in order to possess the concepts we require to understand BP and SS (e.g., *plume*, *bird*, *evolution*, etc.). This entails that experience equips us with the conceptual tools we need to infer SS from BP and thereby gain justification to believe *If BP, then SS* and  $BP > SS$ . One might therefore worry that when we hypothetically abduce SS from BP, our resulting justification derives from experience and, so, can only be a posteriori.

In response, this worry rests on a well-known mistake. While it's true that experience equips us with the conceptual tools we need to acquire justification to believe *If BP, then SS* and  $BP > SS$ , this does not show that such justification even partly derives from experience. All it shows is that experience puts us in a position to acquire such justification by abducting SS from BP. So it remains an open possibility that the justification we then acquire derives solely from rational reflection. That is, even though we need appropriate experiences in order to understand SS and BP, it can still be the case that when we hypothetically abduce SS from BP, we gain a priori justification to believe *If BP, then SS* and  $BP > SS$ .

## 2.2 Justification for BP?

When we hypothetically abduce SS from BP, we obviously draw on the information contained in BP. But, notice, we can only have a posteriori justification to believe BP. BP contains two truths that rational reflection alone is silent about. As such, one might think that when we abduce SS from BP, we therein rely on a posteriori justification to believe BP. But if this is so—and our inference partly rests on such justification—then our inference can only yield

a posteriori justification to believe *If BP, then SS* and  $BP > SS$ .

In response, this line of thought is mistaken about the attitude we must take towards BP in order to *hypothetically* infer SS. Since our inference is hypothetical, it's not based on any belief in BP. Instead, we merely suppose that BP is true and then infer SS on the basis of this supposition. This means that BP is simply a hypothesis that we consider for the purpose of seeing what follows from it. Since BP is merely a supposition, we do not need any justification for it. We don't even need to assign a credence value to BP. All we need to do is understand it, and then reflect on its rational implications. As a result, the fact that BP itself is a posteriori is irrelevant. It does not affect the sort of justification that we acquire when we hypothetically abduce SS from BP.

### **2.3 What About Background Beliefs?**

A third challenge to my argument springs from the fact that, with most abductions, we have to figure the best explanation of a given phenomena in the course of reaching a conclusion. To do this we have to draw on all sorts of background beliefs about, e.g., the range of potential explanations, the nature of explanation, the norms of theory-choice, and the overall coherence of the discipline in question. But at least most of these background beliefs are clearly a posteriori. This is one reason why most of our abductions yield a posteriori justification and knowledge. Most of these inferences rely on our a posteriori justification for the background beliefs that guide our attempts to identify best explanations. Thus, one might worry that the same is true for our hypothetical abduction of SS from BP. But if this is so, then this inference can only yield a posteriori justification to believe *If BP, then SS* and  $BP > SS$ .

In response, I argue that BP's formulation answers this objection. What's relevant is that

BP contains:

2) The best explanation of (1) is that male Kings of Saxony evolved such plumes through a process of sexual selection.<sup>10</sup>

BP therefore tells us what best explains the fact that male Kings of Saxony Birds of Paradise have magnificent brow plumes—viz., sexual selection. So when we suppose that BP is true, we start with this information, and can reason on this basis. The upshot is that when we have BP as a supposition, we do not need to figure out the best explanation of (1) in order to hypothetically abduce SS; we already have this information. For this reason we do not need to draw on our background beliefs about explanation, theory-choice, and the broader discipline in question. The a posteriority of our background beliefs therefore does not infect our reasoning when we hypothetically abduce SS from BP. Free of such a posteriori influence, our inference can thus yield a priori justification to believe *If BP, then SS* and  $BP > SS$ .

## **2.4 Do we Rely on our Past Experience with Successful Abduction?**

Abduction is a fallible process of inference: i.e., even when we abduce in a justified and unobjectionable way, we can get things wrong. So it's possible for justified abduction to lead us into false belief. As a result, one might argue that we need to support our abductions with evidence that they are in fact generally reliable. For since abduction can lead us astray, perhaps we need to have evidence that it typically doesn't in order for our abductions to yield justification.

This suggestion spawns a potential challenge to my argument. This is because it seems like experience is the only thing that can provide evidence about the reliability of abduction. So

---

<sup>10</sup> (1) is: Male King of Saxony Birds of Paradise (whose bodies range from 20-25 cm long) sport magnificent brow plumes that grow up to 50 cm long.

if we need to underwrite our abductions with evidence that they're generally reliable, then it looks like we need to underwrite our abductions with empirical evidence. All of this applies to our hypothetical abduction of SS from BP. If we have to underwrite it with evidence that abduction is reliable, then we have to underwrite it with empirical evidence. But if that's so, then it looks like our inference can only generate a posteriori justification.

I have two points in response. The first is that we may well be able to borrow a response from related debates in epistemology. In particular, it's instructive to note that there's a similar debate in the philosophy of perception: the debate on whether we need evidence that our experiences are reliable in order for them to yield justification for our subsequent beliefs. And in this debate, the dominant (though, of course, disputed) view is that we do not need such evidence. There are two prominent schools of thought here. First, there are those who argue that we do not need evidence that our experiences are reliable because, absent reason to accept sceptical hypotheses on which they are not, we have a “default entitlement” to reject such hypotheses (e.g., Cohen 2000; Burge 2003; Davies 2003; Wright 2004). The idea is that this default entitlement is, itself, enough to support our experiences and ensure they sometimes yield justification. Second, there are those who argue that our experiences sometimes yield justification regardless of whether we have any reason—be it in the form of evidence or entitlement—to think they're reliable (e.g., Pryor 2000, 2004; Peacocke 2004; Silins 2008; Goldman 2008). These writers sometimes express their position by saying that our experiences sometimes yield “immediate justification”. Both of these positions are relevant because, like abduction, perception is fallible. So it seems like both of the above positions naturally extend to abduction. That is, if we have a default entitlement to reject sceptical hypotheses concerning

sensory experience, then why not say the same about abduction? And if our experiences sometimes yield immediate justification, then why not hold that our abductions do too? It's not clear why one would resist these extensions from the case of experience to the case of abduction. As such, I suspect that arguments for the two above views about perception likely also work as arguments for parallel views about abduction. If I'm right, then such arguments also suggest that our abductions can yield justification even when we do not have evidence that they're reliable. Thus, the literature on the epistemology of perception may well provide us with the means to defend the idea that we gain a priori justification for *If BP, then SS* and  $BP > SS$  when we hypothetically abduce SS from BP.<sup>11</sup>

My second point is that, whatever we say about the overall epistemology of abduction, there are clearly some abductions that we do not need to support with evidence that our abductions are generally reliable. This is because, in some cases, we *suppose* that our abductions are reliable, and then infer on this basis. In such cases our supposition automatically underwrites our subsequent abductions with the information that they're reliable. So whenever we suppose that abduction is reliable and then reason on that basis, evidence of abduction's reliability is simply superfluous. Our supposition already does all the work to underwrite our abductions and enable them to yield justification. Importantly, this workaround is readily available in our case. We can simply add a third supposition to BP.<sup>12</sup> One like:

3) Our abductions are generally reliable.

If we make this addition, then when we suppose that BP true for the purpose of thought, we

---

<sup>11</sup> I expand upon and defend this suggestion in other work. This issue is too complex to tackle here.

<sup>12</sup> By packing all the information we might need to infer SS into our suite of suppositions, I adopt a strategy from Chalmers & Jackson (2001) and Chalmers (2012).

automatically underwrite our subsequent abductions with the thesis that abduction is reliable.

This makes empirical evidence of abduction's reliability superfluous, and ensures that when we abduce SS from our supposition that BP, we get justification to believe *If BP, then SS* and  $BP > SS$ . In short, if we carefully specify the content of BP, we can obviate any need to underwrite our hypothetical abduction of SS with evidence that our abductions are reliable. So the fact that such evidence can only be empirical need not infect our reasoning.

## 2.5 Fallibility

As I just discussed, abduction is a fallible process of inference: i.e., even when we abduce in a justified and unobjectionable way, we can get things wrong. It follows that any justification we derive from our abductions is also fallible: viz., it's possible to have this justification for a false belief. All of these points apply when we hypothetically abduce SS from BP. Not only is our hypothetical inference fallible, but so is our consequent justification to believe *If BP, then SS* and  $BP > SS$ . In other words, it's possible to have this justification even when *If BP, then SS* and  $BP > SS$  are false.

This fact suggests an objection to my argument. For one might think—like many important philosophers, e.g., Swinburne 1975; Bonjour 1985; Chishom 1989—that a priori justification is *infallible*; viz., that one cannot have a priori justification for a false belief. But if that's so, then the fallibility of our abductive justification entails that it cannot be a priori.

In response, like many recent authors (e.g., Burge 1993, 1998; Bonjour 1998; Casullo 2003; Warenski 2009) I argue that a priori justification is fallible. There are many cases where people have a priori justification for false beliefs. For instance, suppose that Margot believes that unmarried, adult males are bachelors, and that she bases her belief purely on her

understanding of the concepts involved. But this belief is false. For the Pope is not a bachelor despite being an unmarried adult male. Nonetheless, it's surely true that Margot has some justification for her belief. Her failure to consider the special example of the Pope does not completely undermine her justification. And since this justification derives solely from her competent but perhaps incomplete understanding of *bachelor*, *unmarried*, and other relevant concepts, it's a priori. Margot therefore has a priori justification for a false belief. It follows that a priori justification is fallible. The fallibility of our abductive justification to believe *If BP, then SS* and  $BP > SS$  therefore does not prevent it from being a priori.

## 2.6 Conclusion

In sum, I argue that when we hypothetically abduce SS from BP, we gain a priori justification to believe *If BP, then SS* and  $BP > SS$ . Not only does it appear that such justification derives solely from our reflection on the rational implications of BP, but there's no reason to think that our abduction even partly relies on some a posteriori justification. In light of these findings, I conclude that a priori abduction is not only possible but actual. Our abductive inferences sometimes yield a priori justification for our resulting conclusions.<sup>13</sup> This occurs when we abduce on the basis of suppositions that provide us with all the information we need to rationally draw our inference.

## 3) Thought Experiments

Now that I've argued that a priori abduction is actual, it's time to see what philosophical

---

<sup>13</sup> In other work, I argue that the justification we get when hypothetically abduce SS from BP can well be enough for us to know *If BP, then SS* and  $BP > SS$ . That is, I argue that our hypothetical abduction can provide us with a priori knowledge.



progress we can make with this result. To that end this section develops a new theory of our judgments about philosophical thought experiments (henceforth, I'll usually just refer to them as our judgments). On this theory, we use hypothetical a priori abduction to infer and, absent defeaters, a priori justify such judgments. So, for instance, in regards to Gettier cases, I submit that we hypothetically a priori abduce our judgment about how our imaginary protagonist truly, justifiably believes but does not know.

As a reminder, this theory, if successful, accomplishes four things. First, it defends the reliability and justification of our judgments about thought experiments. Second, it entails that our judgments are a priori justified when justified at all. Third, it reveals that our judgments are the product of a cognitive process that's also at work when we encounter non-imaginary, empirical cases. Finally, this theory empowers us to relinquish the idea that intuitions, *qua sui generis* intellectual seemings, are what occasion and justify our judgments.

### **3.1 The Plan**

To explain how my theory works, I focus on Gettier cases as a paradigmatic example of a thought-experiment. I therefore focus on explaining how we might a priori abduce a judgment about how the protagonist in a Gettier case truly, justifiably believes but does not know. Here's the plan: In the rest of section three I explain how we can hypothetically abduce our Gettier judgment from the suppositions we make to imagine the corresponding case. Then in section four, I argue that this inference yields a priori justification for our Gettier judgment. In section five I argue that the abductive theory I sketch achieves my aim of developing an a priori naturalist theory of the epistemology of thought experiments. Finally, in section six, I spell out some of the implications of the theory I develop.

One important note: I am not going to suggest that we can abduce our Gettier judgments in *exactly* the same way that we can abduce our beliefs *If BP, then SS* and  $BP > SS$ . While I do think that, in both cases, we can perform a priori abductions, the details of each inference turn out to be different. This means that the first half of this paper is not an attempt to model the exact process of a priori abduction that we can also use to infer our judgments. Instead, the first half is an attempt to independently secure the basis of my theory of how we reason about thought experiments: i.e., to independently secure the actuality of a priori abduction.

### **3.2 The Theory: In Brief**

Here, in broad strokes, is how I think we a priori abduce our judgments about Gettier cases. First, we make all the suppositions we need to conceive of the case in question. Then, on the basis of these suppositions alone, we rationally abduce that our protagonist truly and justifiably believes, but does not know. That is, we hypothetically abduce this conclusion from the suppositions that we make to conceive of our Gettier case. This rational inference then provides us with justification for a judgment about how our abduced conclusion is plausible given our suppositions. This means that we gain justification for judgments like:

“If our Gettier case obtains, then our protagonist truly, justifiably believes but does not know”

or

“If our Gettier case were to obtain, then our protagonist would truly, justifiably believe without knowing”

Finally, I'll argue that since our justification for this judgment derives solely from reflection on the content and rational implications of our suppositions, it's a priori. The result is an a priori justified Gettier judgment.

### **3.3 More Details: Imagining the Case**

I'll now detail one way that we can hypothetically abduce a Gettier judgment from the suppositions that we make to conceive of an imaginary Gettier case. Let me emphasize that I develop this as just one way that we might draw such an inference. I suspect that there are multiple ways we can hypothetically abduce our judgment. For present purposes, though, it's enough to simply explain one. This will suffice to flesh out the theory I propose.

We start the process of evaluating a Gettier case by conceiving of the scenario in question. To do this we make a number of suppositions that frame and describe the case. First off, we form suppositions that describe the principal, proprietary details of our scenario. For instance, we might suppose that:

- We have three characters: Smith, Jones, and Brown.
- Smith believes that either Jones drives a Ford or Brown is in Barcelona.
- Smith believes this because he has seen Jones driving a Ford on many recent occasions.
- Smith has no evidence that Brown is in Barcelona.
- By chance, Jones no longer drives a Ford.
- Also by chance, Brown is in Barcelona.

But these six suppositions do not make for a complete thought experiment. When we conceive of a Gettier case, we also fill in the worldly context—the background scenario—that Smith, Jones, and Brown inhabit. And we need to do this. For we need to control the number of variables in our thought experiment. That is, we need to rule out open possibilities that would otherwise paralyze our deliberations about the case. For instance, in our case, if we leave it an open possibility that Smith, Jones and Brown inhabit a world where every true, justified belief magically becomes knowledge (due, say, to the actions of some beneficent God), then it's an

open possibility that Smith magically knows, and we should hesitate to judge that he does not. Similarly, if we leave it an open possibility that our characters inhabit a world where every piece of perceptual evidence runs up against an equal and opposite defeater, then we should hesitate to judge that Smith justifiably believes. But, clearly, whenever we consider a Gettier case like ours, we do not think that we ought to hesitate in such ways. So we must somehow place our case in a worldly context that sets aside these and many other troublesome possibilities.

One straightforward and natural way to do this is to suppose that our Gettier case takes place in a world that's just like how we know the actual world to be, but with the addition that Smith, Jones, and Brown exist, and our Gettier case occurs. In short, one way to proceed is to suppose that our Gettier case occurs in the actual world. I further suggest that, in fact, this is the supposition we typically make when we conceive of a Gettier case. Why is this supposition so effective and apt? Well, for one thing, since our Gettier case is supposed to tell us something about the *actual* nature of knowledge, it seems apt to suppose that this case occurs in the actual world. Another reason this supposition is appropriate is that our Gettier case (like all extant Gettier cases) is not explicitly counterfactual. Yes, we treat it as imaginary, but for all we know and care, it might actually obtain somewhere. Finally, and most importantly, the supposition that our case is actual—which is really a suite of suppositions—is effective at ruling out a host of troublesome, distracting possibilities. For instance, since we know that, in the actual world, many beliefs do not become knowledge, and that perceptual evidence often does not face defeaters, when we suppose that our Gettier case occurs in the actual world, we immediately rule out my two examples of troublesome open possibilities.<sup>14</sup>

---

<sup>14</sup> Do we have to suppose that Gettier cases—and all other thought experiments—occur in the actual world? No. For one thing, I'm sure we can come up with a counterfactual Gettier case. And there are many clearly

If we suppose that Smith, Jones, and Brown inhabit the actual world, we thereby draw on our knowledge of the actual world to flesh out our hypothetical scenario. Our knowledge dictates what we suppose. So, for instance, since we know that abduction and perception are both generally reliable, when we suppose that our characters inhabit the actual world, we thereby suppose that in our hypothetical scenario, abduction and perception are generally reliable. Even more importantly, when we we suppose that our Gettier case occurs in the actual world, we also thereby suppose that their scenario features the same paradigm instances of knowledge that we know to obtain in the actual world. These suppositions enable us to compare many paradigm instances of knowledge with Smith's belief. That is, once we suppose that certain paradigms of knowledge obtain in our imaginary case, we can think about whether, in the world of our case, Smith's belief resembles or fails to resemble such paradigms.

When we run this comparison, we can immediately see that Smith's belief compares unfavourably. This is because our paradigm cases of knowledge are beliefs that, in addition to being true and justified, are characteristically: i) causally connected to the facts, ii) produced by a reliable process, iii) not the product of a false lemma, and/or iv) safe (i.e., such that in all nearby worlds where the subject in question holds the same belief, this belief is true). In short, our

---

counterfactual thought-experiments—such as twin-earth cases. With all such cases, we clearly do not suppose that they occur in the actual world. But this difference has no effect on my account. For with such cases, we instead suppose that they occur in a carefully selected counterfactual world. And this sort of supposition is equally good for my purposes. For my main point here is that we always locate our specific thought-examples within a worldly context—be it actual or counterfactual—in order to rule out troublesome possibilities and, thus, infer definitive conclusions. We find evidence for this practice in the fact that when we suppose that a twin-earth case occurs in a certain counterfactual scenario, we always take care to carefully describe this world and thereby rule out troublesome possibilities. Usually we do this by contrasting and comparing our counterfactual scenario to the actual world (e.g., we might say “It's just like the actual world except for the fact that XYZ, not H2O, fills its lakes, rivers, and oceans”). This lets us make it clear that our test-case occurs in a counterfactual world, and yet closely describe this world by reference to the actual world we know so well. We thus make all sorts of background suppositions about our counterfactual scenario, and these suppositions not only rule out troublesome possibilities. My argument is that these background suppositions are what allow us to hypothetically abduce our judgments about thought experiments from the just the suppositions we make to conceive of such cases.

paradigm cases of knowledge are beliefs that look *non-accidentally true*. Given the setup of our imagined case, none of this holds for Smith's belief. His belief: i) fails to be causally connected to the facts, ii) is produced via a false lemma, iii) is produced by an unreliable process (since it's produced via false lemma), and iv) is plausibly unsafe. In short, his belief looks to be *accidentally true*.

### 3.4 More Details: Drawing the Inference

I'll now explain how we might draw on all of our suppositions to hypothetically abduce our way to a judgment about how Smith truly, justifiably believes but does not know.

First, we can hypothetically abduce that Smith truly and justifiably believes from our suppositions: i) Smith believes that either Jones drives a Ford or Brown is in Barcelona, ii) Smith believes this because he has seen Jones driving a Ford on many recent occasions, iii) Brown is in Barcelona, and iv) perception is a reliable source of justification. For we can first deduce from (i) and (iii) that Smith's belief is true, and then draw on suppositions (i), (ii) and (iv) to abduce that Smith's belief is also justified. We build the last supposition into our imaginary case when we suppose that it occurs in the actual world. For since we know that, in the actual world, perception is a reliable source of justification, we therefore suppose that this is also true in the world of our thought experiment.

Second, I argue that we can hypothetically abduce that Smith fails to know from the suppositions we make to conceive of our Gettier case. To do this we can use our finding that Smith's belief compares unfavourably to the paradigm cases of knowledge that we've built into our thought-example.

To explain how we can we perform this hypothetical inference, let's first think about how

we'd reason if we were to encounter Smith not in an imaginary case, but in real life. In particular, what would we judge to be the best explanation of the discrepancy between his belief and our paradigms of knowledge? It seems like we could justifiably conclude that the best explanation of this discrepancy is that Smith's belief fails to be knowledge. This is because, first and foremost, the theory that Smith lacks knowledge would effectively explain the discrepancy in question. For if his belief fails to be knowledge, then it makes sense that his belief is so unlike our paradigms of knowledge.

The theory that Smith lacks knowledge would also fit well with our norms of explanation and theory-choice. For instance, this theory would seem to be the simplest available explanation of the discrepancy between Smith's belief and our paradigms of knowledge. This is because in order to explain this discrepancy while holding that Smith's belief amounts to knowledge, we'd need some account of how and why instances of knowledge can be so unlike our present paradigms of such. And we'd likely also have to re-evaluate and recast our paradigms to better reflect the new diversity among instances of knowledge. In contrast, we wouldn't need any of these complications on the theory that Smith lacks knowledge. We'd be able to explain the discrepancy in question while retaining our present paradigms, and we would not need an account of how instances of knowledge can be so unlike these paradigms. Overall, then, it seems that if we were to encounter a real-life version of our Gettier case, we could justifiably conclude that since the theory that Smith lacks knowledge is simple, effective, and so forth, it's the best explanation of the discrepancy we note between Smith's belief and our paradigms of knowledge.

I bring this up because I argue that we can reason in almost exactly the same way about an imaginary version of our Gettier case. The only difference is that when our case is

imaginary, it's the suppositions we make to conceive of it that provide us with the information we use to identify best explanations in the world of this case. That is, when our case is imaginary, we reason about what, given our suppositions, best explains the discrepancy between Smith's belief and our supposed paradigms. Nonetheless, our suppositions are such that we can reason in the much the same way that we could if we were to encounter a real-life version of our case. This is because when we conceive of our imaginary case, we suppose that it occurs in the actual world. So our suppositions about the world of our case mirror our knowledge of the actual world. This means, for one thing, that we suppose that our case occurs in a context with the same paradigms of knowledge that we know to obtain in the actual world. It also means that we suppose our case occurs in a context with the same norms of explanation and theory-choice that we know to obtain in the actual world. In short, our suppositions provide us with all the information that would, in the form of knowledge/belief, drive our reasoning about a real-life Gettier case. As such, we can, just on the basis of our suppositions, justifiably conclude that in the world of our thought-experiment, the theory that Smith lacks knowledge is the best explanation of the discrepancy between his belief and the paradigms of knowledge that we build into our case.

There's now a straightforward path for us to hypothetically abduce that Smith lacks knowledge from the suppositions we make to conceive of our Gettier case. For once we justifiably conclude, on the basis of our suppositions, that the relevant discrepancy in our case is best explained by the theory that Smith lacks knowledge, we can then infer to this best explanation. That is, we can justifiably and hypothetically abduce from our suppositions that, in the world of our imaginary Gettier case, Smith lacks knowledge.



To complete our inference, all that remains is for us to combine our conclusions: i) that Smith truly and justifiably believes, and ii) that he fails to know. This leaves us with the final, hypothetically abduced conclusion that Smith truly, justifiably believes but does not know.

### 3.5 The Contents of our Judgments

While I certainly think that we can hypothetically abduce that Smith truly, justifiably believes but does not know from the suppositions we make to conceive of our Gettier case, I *do not* think that our eventual judgment about this thought experiment has the content:

“Smith truly, justifiably believes but does not know”

That is, I do not think that the conclusion we hypothetically abduce from our suppositions is identical to the final judgment we form about our Gettier case. This is because, for one thing, we know that this case is imaginary, so we surely do not judge that Smith exists and has beliefs. What's more, we know that our *hypothetical* abduction does not provide us with justification to judge that Smith truly, justifiably believes but does not know. For in order to get such justification, we'd need to have justification for the premises we infer this conclusion from. But in this instance, we obviously do not have such justification, as all of our premises are mere suppositions—the suppositions that detail our imaginary case.

So what then is the content of our judgment? I hold that our judgment takes one of the following contents:

“If our Gettier case obtains, then Smith truly and justifiably believes but does not know.”  
or

“If our Gettier case were to obtain, then Smith would truly and justifiably believe without knowing.”

or perhaps even:

“In the fiction of our Gettier case, Smith truly and justifiably believes but does not know.”

That is, I propose that our judgment about this case—and any other Gettier case—is a conditional, counterfactual, or 'meta-fictional' claim about what's true in the fiction we create for our thought experiment. This is because when we hypothetically abduce that Smith truly, justifiably believes but does not know from the suppositions that detail our Gettier case, what we gain justification to judge is something about how our abduced conclusion is plausible given the suppositions that detail our case. Conditional, counterfactual, and meta-fictional judgments naturally express this sort of content.

By offering this account of the content of our Gettier judgments, I invite controversy, for there's plenty of debate about their exact content (see: e.g., Williamson 2007, Ichikawa & Jarvis 2009; Ichikawa 2009; Malmgren 2011). However, I won't adjudicate this complex debate here, so a sustained defense of my suggested account will have to wait (I pursue it in other work). Nonetheless, I think that, regardless of whether we hypothetically abduce our judgments, it's *prima facie* plausible that our judgments about Gettier cases—and other thought experiments—are conditionals, counterfactual, or meta-fictionals. This is because since we know that we're reasoning about an imaginary test case, it seems apt for our judgments to express how certain conclusions are plausible given the suppositions we make to imagine this case. I should also point that there are other people who argue—for completely independent reasons—that our judgments take the contents I suggest. For instance, Williamson (2007) argues that our judgments are counterfactuals, and Ichikawa and Jarvis (2009) argue that our judgments are meta-fictionals.

#### **4) A Priori Justification**

I argue that when we hypothetically abduce our way to an judgment like “If our Gettier case obtains, then Smith truly and justifiably believes but does not know”, this judgment is a priori justified (absent defeaters). The core of my case is simple: we gain justification for this judgment by hypothetically abducting a conclusion from just the suppositions that we make to conceive of our Gettier case. For these suppositions provide us with all the information we need to: i) reason that the discrepancy between Smith's belief and our supposed paradigms of knowledge is best explained by the theory that Smith fails to know, and then ii) subsequently abduce our judgment. So, in broad terms, we gain justification for our judgment by simply reflecting on the rational implications of the suppositions we make to conceive of our case. As such, it seems that we derive our justification from rational reflection alone. Conclusion: our Gettier judgment is a priori.

##### **4.1 Don't We Rely On A Posteriori Knowledge About the Actual World?**

To further press the above argument, I'll now consider and rebut some notable reasons for thinking that our Gettier judgment is a posteriori justified if justified at all.

One such reason centers on my claim that when we consider our imaginary Gettier case, we suppose that this scenario occurs in the actual world. For it follows from that we use our knowledge of the actual world to inform the suppositions we draw when we conceive of our Gettier case. In other words, our worldly knowledge dictates the suppositions we draw. But our worldly knowledge is almost entirely a posteriori. In particular, our knowledge of paradigm instances of knowledge is certainly a posteriori. So one might worry that when we abduce that

Smith truly and justifiably believes but does not know we somehow rely on the a posteriori justification we have for the knowledge that informs our suppositions. But if this is so, and our reasoning rests on such justification, then our inference can only yield a posteriori justification for our judgment.

In response, this objection misses two important points. The first is that, on my view, we abduce that Smith truly and justifiably believes but does not know solely from the suppositions we make to specify our imaginary Gettier case. As such, none of our worldly knowledge serves as a premise in our abductive inference. Our framing suppositions provide us with all the premises we need. The second key point is that our knowledge of the actual world merely informs our suppositions—i.e., dictates their content—it does not justify or support them. After all, how could it? Suppositions are not the sort of things that have or need justification.<sup>15</sup> When we keep these two points in mind, we see that while our knowledge of the actual world is a posteriori, this knowledge does not play a justificatory role when we abduce that Smith truly and justifiably believes but does not know. This knowledge merely tells us what to suppose when we imagine our Gettier case. And our resulting suppositions then provide us with all the information we need to perform the inference in question. As a result, while our a posteriori knowledge of the actual world importantly guides us in some sense when we infer that Smith truly and justifiably believes but does not know, it does not justify us, and there is no reason to think that our inference therefore yields only a posteriori justification for our judgments.

## **4.2 What about Background Beliefs?**

---

<sup>15</sup> Here I'm talking about epistemic justification. Suppositions are not epistemically justified or unjustified in the way that beliefs are. To be sure, though, suppositions can be justified in various pragmatic senses. For instance, a supposition is justified in one sense to the extent that it fits the purpose or plot of its ensconcing thought-example.

A second—and now familiar—challenge to my argument springs from the fact that, with most abductions, we have to figure the best explanation of a given phenomena in the course of reaching a conclusion. To do this we have to draw on all sorts of background beliefs about, e.g., the range of potential explanations, the nature of explanation, the norms of theory-choice, and the overall coherence of the discipline in question. One might therefore suspect that when we abduce that Smith truly and justifiably believes but does not know, we also rely on such background beliefs. For instance, one might think that we rely on such beliefs to figure out that Smith's failure to know best explains, in our imagined case, the discrepancy between his belief and our paradigms of knowledge. But if so, then this is a problem for my argument, for at least most of our background beliefs are a posteriori. So if they play the suggested role, then our abduction can only yield a posteriori justification for our resulting judgments. This is because now our inference relies on our a posteriori justification for the background beliefs that, according to this objection, guide our attempt to identify the best explanation of the discrepancy between Smith's belief and our supposed paradigms of knowledge.

In response, remember that I propose that when we conceive of our Gettier case, we suppose that this case occurs in a world that's just like the actual world except for the addition that our case occurs. By doing this, we therefore suppose that our scenario features the same norms of theory-choice, candidate explanations, standards of explanation, etc., that we know obtain in the actual world. This means that when we conceive of a Gettier case, we provide ourselves with suppositions that carry all the background information we need to figure out the best explanation of the discrepancy between Smith's belief and our paradigms of knowledge. That is, when conceive of our Gettier case, we supply ourselves with suppositions that can play

the best-explanation-identifying role often filled by our background beliefs. So we can identify best explanations (such as Smith's failure to know) in our hypothetical scenario without drawing on our actual background beliefs. The result is that these beliefs do not play any role in our abduction. Their a posteriority therefore does not infect the justification that our hypothetical inference supplies to our judgments.

### 4.3 Do we Rely on Evidence that Abduction is Reliable?

As I mentioned earlier, someone might argue that since abduction is fallible, we must have evidence that our abductions are generally reliable in order for them to yield justification. If this claim is true, it prompts a familiar objection to my theory. For we can only gain empirical evidence of abduction's reliability. Thus, if we have to support each of our abductive inferences with evidence that our abductions are generally reliable, then we have to support each of our abductions with empirical evidence. But it's that's so, then our abductions—including our hypothetical abduction to a Gettier judgment—can only yield a posteriori justification.

When I discussed this objection earlier in the paper, I argued that at the very least, there are some abductions that we do not need to underwrite with empirical evidence of abduction's reliability. This is because we can suppose that abduction is reliable, and then reason on that basis. In such instances, our supposition underwrites our subsequent abductions with the information that they're reliable, and renders evidence of such superfluous. I then suggested that we could use this loophole to ensure that when we hypothetically abduce SS from BP, we get a priori justification for *If BP, then SS* and  $BP > SS$ —all we have to do is add the supposition “abduction is generally reliable” to BP.

Does the same loophole apply when we hypothetically abduce “Smith truly, justifiably

believes but does not know” from the suppositions we make to conceive of our Gettier case? Yes it does. But this time we don't even have to add a new supposition to our inference base. This is because, in the present example, when we conceive of our Gettier case, we suppose that it occurs in a world that's just like the actual world except for the addition that our case occurs. But we know that in the actual world, abduction is generally reliable. Thus, when we conceive of our Gettier case, we suppose that it occurs in a world where abduction is reliable. That is, when we reason about the world of our imaginary Gettier case, we reason on the basis of the supposition that abduction is reliable. So if we need to (though, as I argued in 2.4, there's good reason to think we do not), we can rely on this supposition to underwrite our abductions about our case, and thus obviate any need for evidence of abduction's reliability. Therefore, the empirical nature of such evidence does not threaten the a priority of any justification that we obtain when we hypothetically abduce that Smith truly, justifiably believes but does not know.

## **5) A General Theory**

To review: I submit that when we consider a Gettier case, we can a priori abduce our way to judgments like:

“If our Gettier case were to obtain, then Smith would truly and justifiably believe without knowing.”

I further suggest that we can employ the same pattern of inference to form and justify our judgments about any other philosophical thought experiment; be it a twin-earth, Frankfurt, or trolley case. In each instance, we can conceive of our imagined scenario by way of a host of suppositions, and then hypothetically abduce conclusions about this scenario from, and on the

basis of, these suppositions. Our inference then yields, if all goes well, a priori justification for conditional, counterfactual, or meta-fictional judgments expressing how our abduced conclusions are plausible given the suppositional setup of our thought experiment. In saying this, I completely accept that there is bound to be plenty of variation in how we reason through different thought experiments. I just suggest that such differences are differences in suppositions and conclusions, not differences of inferential process.

Now that we have both an overall picture of the theory I suggest, and a detailed instance of how it might work in practice, let's think about whether this abductive theory actually achieves the four goals that initially set for myself. Does this theory: i) defend the reliability and justification of our judgments about thought experiments, ii) make it plausible that our judgments are a priori justified when justified at all, iii) reveal that our judgments are the product of a cognitive process that's also at work when we encounter non-imaginary, empirical cases, and iv) empower us to relinquish the controversial idea that intuitions, *qua sui generis* intellectual seemings, are what occasion and justify our judgments?

Two answers are already clear. First, I've already argued at length that if we hypothetically abduce our judgments in the way I propose, then they're a priori justified when justified at all. Second, it's evident that if the theory I suggest is successful, then we can dispense with intuitions in our epistemology of thought experiments. For intuitions play neither a psychological nor evidential role in this theory's account of we form and justify our judgements.

However, it's not immediately whether the theory I suggest satisfies the first and third desiderata from above. So I'll now argue that it does.

### **5.1 Imaginary and Non-Imaginary Experiments**



Does the theory I suggest entail that our judgments are the product of a cognitive process that's also at work when we encounter non-imaginary, empirical cases? It clearly does. For we constantly use abduction to make sense of the empirical data we gather about the non-imaginary experiments and scenarios we encounter in science and everyday life. Of course, our abductions are not hypothetical when they operate on the basis of empirical data instead of suppositions, but, despite this difference, we're still left with an important similarity between our evaluations of thought and non-thought experiments. In both cases we confront some 'data'—be it suppositional or empirical—and we infer conclusions by figuring out what best explains this data.

## **5.2 Reliability and Justification**

Does the theory I suggest make it plausible that our judgments about thought experiments are generally reliable and justified? I think it does. First off, as I argued in section 1, our everyday practice of hypothetical reasoning makes it clear that when we rationally infer a conclusion from a set of suppositions, we gain justification, absent defeaters, to believe conditionals and counterfactuals with our suppositions in antecedent position and our conclusion in consequent. Thus, there's good reason to think that whenever we rationally abduce a conclusion from the suppositions to conceive of a thought experiment, we gain justification, absent defeaters, to form judgments like:

“If our Gettier case obtains, then Smith truly and justifiably believes but does not know.”  
or  
“If our Gettier case were to obtain, then Smith would truly and justifiably believe without knowing.”

The theory I suggest also supports the reliability of our judgments. This is because

abduction is a generally reliable process of inference. So if we use abduction to form our judgments, then we employ a generally reliable belief forming process. One way to make the reliability of abduction very evident is to point out that abduction had better be reliable, or widespread scepticism looms. This is because we use abduction all the time: it's what we primarily employ in both everyday and scientific reasoning. For instance, my belief that my alarm will go off at 9:30 AM tomorrow morning is the product of abduction—in particular, induction—on past experience. And when physicists concluded that supernovae create heavy elements, they used abduction. For their reasoning was that supernovae are the only observed phenomena with the heat and pressure required to forge heavy elements.

The ubiquity of abduction is such that if it's unreliable, a great many of our everyday and scientific beliefs are poorly founded, and likely do not count as knowledge. But it's clearly not the case that most of our everyday and scientific beliefs fail to be knowledge. I know that my alarm will go off at 9:30, and physicists know that supernovae create heavy elements. So on pain of rampant scepticism, it must be the case that abductive reasoning is a generally reliable cognitive process.<sup>16</sup> Therefore, on the theory I suggest, it must also be the case that our judgments about thought experiments are generally reliable.

## **6) Implications**

### **6.1 Synthetic A Priori Knowledge**

Since Kant drew a distinction between the analytic and the synthetic, many people have

---

<sup>16</sup> While I doubt that hypothetical abductive reasoning is a distinct natural kind of cognitive process, the same reasoning applies to this more specific 'type' of reasoning. That is, here it's also clear that, on pain of rampant scepticism, hypothetical abduction is reliable. For we use hypothetical abduction all the time in everyday reasoning and scientific enquiry. This is shown by my examples involving commonplace conditional belief about Sergei.

thought that we have some synthetic a priori knowledge. Many people have also thought that philosophy is unique in being the study of synthetic a priori truths. But this position has always faced the problem of explaining how we might gain a priori justification to believe synthetic truths. This problem has not faded with the years. Though Kant's work is over two centuries old, we still grapple with the problem of synthetic a priori knowledge.

I suggest that a priori abduction is a ready answer to our problem. For once we accept the actuality of a priori abduction, we provide ourselves with a good account of how it is that we can acquire synthetic a priori knowledge. Namely: we a priori abduce our way to such knowledge.

This abductive account has two laudable features. First, it does not require us to posit the existence of controversial things like intuitions—*qua* intellectual seemings. All this account requires is that we have the capacity to perform abductive inferences. And this is a capacity that we clearly have.

Second, this abductive account of the synthetic a priori readily supports the Kantian position that philosophical knowledge is often synthetic a priori. This is because a significant portion of our philosophical knowledge (assuming there is any) derives from the synthetic judgments we form in our evaluation of thought experiments. But as we've just seen, there's reason to think that we use a priori abduction to infer and a priori justify such judgments. So on this theory, any philosophical knowledge that our judgments manifest or entail—i.e., a lot of philosophical knowledge—is synthetic a priori. The overall result: if we acknowledge that a priori abduction is actual, we can say that philosophy is a discipline of the synthetic a priori—just as Kant envisioned—and yet explain this with just our evident ability to perform abductions.

## 6.2 Philosophy *Sans* Intuitions

I further suggest that a priori abduction offers us the prospect of completely dispensing with intuitions as a source of justification in epistemology. At the moment, intuitions feature in many accounts of philosophical knowledge, justification, and methodology. Theorists cite intuitions in order to explain and/or defend things like: synthetic a priori knowledge, our judgments about thought experiments, foundationalism, mathematical knowledge, moral knowledge, modal knowledge, and more. But the idea that intuitions exist and have key justificatory force is deeply controversial. Some people even argue that we lack good reason to think that intuitions exist at all (e.g., Hintikka 1999, Cappelen 2012). So while intuitions might well be put to good philosophical work, they're problematic as a basis for philosophy.

We should therefore have interest in any opportunity to defend and explain the things we want in a way that does not require intuitions. I think that a priori abduction offers us just this sort of opportunity.<sup>17</sup> For I suspect that a priori abductions can play all of the roles that we presently attribute to intuitions. In this paper I've already begun to explain how this might work. Not only do I here suggest that our capacity for a priori abduction can explain how we form and a priori justify our judgments about philosophical thought experiments, I also suggest that a priori abduction can serve as a source of synthetic a priori knowledge. I thereby argue that we can fill two of the primary roles for intuitions with a priori abduction. It remains to be seen whether we can extend these results and ideas to explain and/or defend things like moral and mathematical knowledge, but I hope we can. If my hope succeeds, then our newly recognized

---

<sup>17</sup> We could, of course, re-conceive of intuitions as the conclusions of hypothetical abductions. I'm completely open to that option. I do not suggest that we dispense with use of the term "intuition" altogether. Just with the idea that intuitions play a key role in epistemology.

capacity for a priori abduction may well allow us to eventually dispense with intuitions altogether. I, for one, welcome that prospect.

## **Chapter 2: Inference and the Epistemology of Grounding**

Metaphysicians have recently embraced the notion of metaphysical grounding. That is, to quote Kit Fine, many people now accept that there is:

...a distinctive kind of metaphysical explanation, in which explanans and explanandum are connected, not through some sort of causal mechanism, but through some constitutive form of determination (Fine, 2012a).

This view entails that reality has metaphysical structure, as its various constituents fall into a hierarchy of those that ground and those that are grounded. Proponents of grounding argue that one task—if not the primary task—of metaphysics is to understand this structure by understanding the grounding relationships inherent in reality. That is, it's now widely held that metaphysics is in part the study of “what grounds what” (Schaffer 2009).

In this new zeitgeist it's commonplace to make claims like “non-moral truths ground all moral truths”, “the fact that the table is white and square is grounded in the fact that the table is white, and the fact that the table is square”, and “the water's phase is grounded in the kinetic energy of its component molecules”. In fact, metaphysics seems to be in the midst of a important transition, as talk of grounding begins to replace long-dominant modal discourse. For it seems that many metaphysicians now think that grounding relations are more explanatory and natural

than modal and mereological relations. Accordingly, the literature is full of debate about grounding and its import. On one side, there is a healthy debate about what grounds what. Much of this debate concerns the relata of grounding relations. Some writers think that facts are the only relata of grounding relations (e.g., Fine, 2012a; Sider 2011), whereas others hold that objects, properties, and/or states of affairs can also stand in such relations (e.g., Schaffer 2009). On other flanks, several writers have begun to explore the meta-theory of grounding, as the last couple of years have seen a lot of work on, e.g., the logic of grounding, the fundamentality of grounding, and the semantics of “grounding” talk (e.g., Schaffer 2009, 2012; Sider 2011; Fine 2012a, 2012b; Bennett 2011, MS; Williams 2012).

Despite all of this thought and progress, there is one field that has seen relatively little attention: namely, the epistemology of grounding. Here we face questions like: how might we come to know what grounds what? And how do we determine whether something is grounded or ungrounded? Little work has been done on these questions, and we do not have any widely-acknowledged answers. This puts us in an awkward position. On the one hand, we're starting to make all sorts of claims about the presence of grounding relations, and it seems like many of these claims are justified and even count as knowledge. But on the other hand, it's not clear how we might test, gain evidence for, and generally come to know such claims. In short, we face a significant theoretical deficit in the epistemology of grounding and grounding-centric metaphysics.

One way to start filling this deficit is to see whether we can gain any inspiration from existing debates on nearby topics. For instance, themes from modal epistemology might well

illuminate the epistemology of grounding. And so might the epistemology of counterfactuals.<sup>18</sup> However, in this paper, I want to focus on a debate in the philosophy of mind that turns out to be very relevant to the epistemology of grounding. This debate comes in a couple of guises. Some cast it as the debate on whether “reductive explanation requires a priori entailment” (Chalmers & Jackson 2001. Also: Levine 1993; Block & Stalnaker 1999), whereas others talk about whether metaphysical “determination” requires “a priori determination” (Jackson 2006). On either guise, the issue here is whether one truth,  $X$ , reductively explains/determines another truth,  $Y$ , only if we (or at least idealized agents) can infer  $Y$  from  $X$  and thus know the material conditional  $X \supset Y$  a priori.<sup>19</sup> On one side, people like Lewis (1994), Jackson (2001, 2003, 2006, 2010) and Chalmers (2001, 2009, 2012) argue that, yes, reductive explanation requires inferrability and a priori entailment. And in the opposing camp, we have people like Block & Stalnaker (1999), Byrne (1999), Tye (2009) and Levine (2010).

This debate is relevant because it seems natural to read “reductively explains” and “determines” as predecessors, analogues, or even extensional twins of “grounds”. For, note, grounding relations are typically taken to be or provide (to return to our opening quote) “a type of metaphysical *explanation*” and “some *constitutive* form of *determination*”. In any case, even if we don't assimilate “reductively explains” and “determines” to “grounds”, the *prima facie* similarity of these terms is enough in this context, for it suggests that, given the debate on

---

<sup>18</sup> In other work, I discuss whether i) modal epistemology, and ii) the epistemology of counterfactuals can teach us about the epistemology of grounding.

<sup>19</sup> How do we get from the claim that we can infer  $B$  from  $A$  to the claim that we can know  $A \supset B$  a priori? Well, according to Chalmers & Jackson (2001), Jackson (2006), and Chalmers (2012) the answer starts with the claim that if we can infer  $B$  from  $A$ , then we can do so even when we treat  $A$  as merely a hypothesis. That is, we can suppose that  $A$ , and then infer  $B$  on the basis of our supposition. And when we do so, we can then discharge our supposition and justifiably conclude that  $A \supset B$ . The crucial idea is that this sort of hypothetical inference yields a priori justification for our resulting belief in  $A \supset B$ . See Chalmers & Jackson 2001 for a longer explanation.

whether reductive explanation requires inferrability, we should think seriously about whether grounding requires the same. That is, does X ground Y only if we can infer Y from X and thus know  $X \supset Y$  a priori?<sup>20</sup>

This question is important because it has big implications for the epistemology of grounding and the methodology of grounding-centric metaphysics. For if we find that, yes, grounding requires inferrability, then we're presented with a powerful way to gain evidence and knowledge of grounding relations. This is because inferentialism entails that we can test whether A grounds B just by determining whether we can infer B from A. In short, what's at stake here is whether inference relations are a guide to grounding relations.

In this paper I try to answer the question of whether grounding requires inferrability and a priori entailment. To this end, I evaluate the plausibility of a “yes” answer; viz., I evaluate the view I call:

**Inferentialism:** X (wholly) grounds Y only if we (or at least idealized agents) can infer Y from X.<sup>21</sup>

---

<sup>20</sup> If we think of properties and objects—instead of truths or facts—as the relata of grounding relations, then our question is, roughly: does X ground Y only if we (or at least idealized agents) can infer Y-truths from X-truths?

<sup>21</sup> 1) Whenever I talk about whether we can infer Y from X, I'm really talking about whether it's possible for at least **idealized agents** to infer Y from X. This is because contingent cognitive and environmental limitations can easily prevent non-idealized subjects from performing perfectly good inferences. We therefore need to consider subjects who are idealized in that they use error-free, comprehensive reasoning, and enjoy ample time and mental energy. Only then can we be confident that any inferential failure is not merely the result of inadequate reasoning, sloppiness, or fatigue.

This qualification is taken for granted across the literature on reductive explanation, so I won't call much attention to it (see: Block & Stalnaker 1999, Chalmers & Jackson 2001, Levine 2010, Chalmers 2012). As such, while I will sometimes remind readers of this qualification, I will often elide it.

2) When I talk about whether we (or at least idealized versions of us) can infer Ms, I'm talking about whether we can *justifiably* draw such inferences. That is, the issue is whether we can infer macrophysical truths from P in a way that has epistemic warrant and (at least typically) generates knowledge.

3) As I note above, it's important to see that whenever I talk about grounding, I'm talking about **whole** or complete grounding, not partial grounding. Roughly, X wholly grounds Y iff X is the complete and total ground of Y—i.e., iff X grounds Y without the 'help' of any other factor. X partly grounds Y iff X does something to ground Y, but Y is grounded in more than X. Henceforth, I will often drop “wholly” for the sake of simplicity.



To do so I consider arguments both for and against this theory. But in the end, I argue that Inferentialism is false, as it faces powerful counterexamples. This means that I ultimately argue that inference relations are not a guide to grounding relations, and that we cannot learn much about the structure of reality by just exploring the limits of inference.

Here's the plan for the paper. In section 1 I further explore inferentialism's implications for both the epistemology of metaphysics and other areas of philosophy. In section 2 I highlight thinkers who argue that reductive explanation requires a priori entailment. I also point out that some of these thinkers have recently argued for new conclusions that approach (or even amount to) inferentialism. In section 3 I explore the positive case for inferentialism and suggest that it's actually quite powerful. This is because: i) there is a *prima facie* plausible general argument for inferentialism, and ii) there are no easy counterexamples. Finally, in sections 4 and 5 I develop my argument against inferentialism. I focus on trying to establish that we (not even idealized versions of us) cannot infer macrophysical truths from the complete microphysical truth. I do this because since (as virtually every proponent of grounding agrees) the complete microphysical truth wholly grounds macrophysical truths, inferentialism entails that we can infer the latter from the former.<sup>22</sup> So when I argue that we cannot infer macrophysical truths from the microphysical truth, I also argue that inferentialism is false.

## 1) The Importance of Inferentialism

---

4) As you can see, I want to de-emphasize the issue of a priori entailment. For I think that the core issue here is whether grounding entails inferrability. We don't need to tangle with the additional issue of whether hypothetical inference from X (ground) to Y (grounded) generates a priori justification to believe that  $X \supset Y$ .

<sup>22</sup> The only potential dissidents here are those who argue for priority monism (e.g., Schaffer 2010).

## 1.1 Metaphysics

It should already be clear that inferentialism offers us a way to learn about the structure of reality. For if inferrability is a necessary condition upon grounding, then we can test whether X grounds Y by determining whether we can infer Y from X. On one hand, if we find that we cannot so infer Y, then we can immediately conclude that X does not ground Y. And on the other hand, if we find that we can infer Y from X, then this is some evidence that X grounds Y. Inferentialism therefore offers a method to not only test grounding claims for failure, but also to gain evidence for them.

Inferentialism also presents us with a way to investigate what properties and truths are fundamental. We just need one additional—and plausible—assumption: that all non-fundamental truths are grounded in the fundamental truth. This lets us use the inferentialism to assess suggestions about the fundamental. All we need to ascertain is whether we (or at least idealized versions of us) can infer non-fundamental truths from the candidate fundamental truth. If not, then inferentialism tells us that our candidate fails to ground all non-fundamental truths. And given our additional assumption, it then follows that our candidate cannot be the fundamental truth after all. Inferentialism therefore allows us to test any suggestion about the fundamental for failure. But this is not all, for the method in question might also be useful to identify the correct account of the fundamental. If only one otherwise plausible account passes the inferentialist test, then this test reveals the fundamental *via negativa*.

## 1.2 Other Implications

With its clear import in the epistemology of grounding and the methodology of metaphysics, inferentialism can well be put to work in other areas of philosophy. For the issue of

what-grounds-what is—in the new grounding-centric zeitgeist—central to meta-ethics, philosophy of mind, philosophy of language, and many other disciplines. In meta-ethics, we have the issue of whether non-moral truths ground moral truths. In philosophy of mind, there's the question of whether physical truths ground phenomenal truths. And in philosophy of language, there's the question of whether non-semantic truths ground semantic truths. In each case inferentialism offers a way to test for the grounding relation at issue. In doing so, inferentialism also suggests a way to argue against each of these grounding claims. For if inferentialism is true, then one can undermine such grounding claims by arguing that, e.g., semantic truths cannot be inferred from non-semantic truths.

This type of argument is familiar in meta-ethics. This is because most philosophers agree that we (even idealized versions of us) cannot successfully infer moral truths from non-moral truths. Or, in more familiar terms, most people accept that there's an is-ought gap. As such, inferentialism gives rise to an argument for the conclusion that non-moral truths fail to ground moral truths. This pattern of argument exists in the literature. For example, Blackburn (1971, 1985) criticizes realist cognitivism in much this way (though he speaks in terms of “supervenience” instead of “grounding”), and Jackson (1998, 2003) uses this sort of argument to contend that realist cognitivists—like him—need to accept that moral truths can be inferred from an appropriate collection of non-moral truths (in other words, Jackson argues that realist cognitivists must deny that there's an is-ought gap).

A similar argument against physicalism is familiar from the philosophy of mind. This argument begins with the plausible contention that we cannot infer phenomenal truths from any collection of physical truths. This claim gains support from thought-experiments like the famous

case of Mary and her black-and-white room. In this example, it's supposed to be intuitive that Mary, trapped in her room: i) can learn all of the physical truths from books, TV, and conversation, but ii) cannot learn phenomenal truths about what it's like to experience colour until she escapes and actually experiences colour. If both these claims are true, then Mary cannot infer such phenomenal truths from the physical truth. But Mary is an ideal reasoner, so if she cannot infer such phenomenal truths, then they cannot be thus inferred. Whatever one thinks about Mary's situation, it's clearly relevant to the anti-physicalist argument on offer. For if phenomenal truths cannot be inferred from the physical truth, then inferentialism dictates that the latter does not ground the former. And if physical truths do not ground phenomenal truths, physicalism is false. Inferentialism therefore offers a straightforward way to use Mary's case to attack physicalism. This basic pattern of argument also exists in the literature. For instance, see Alter (2007, forthcoming) and Chalmers (2004, 2009).

## **2) The Precedent for Inferentialism**

Several writers argue that reductive explanation and determination require inferrability (and, so, a priori entailment). And since it seems apt to translate “reductive explanation” & “determination” as “ground”, it's also seems apt to interpret such writers as effectively arguing for inferentialism.

The first flickers of this line of thought occur in the work of pioneering physicalists such as Smart (1959), Armstrong (1968) and Lewis (1972). For they argue that in order to reduce mental (or moral, epistemic, etc.) phenomena to fundamental physical phenomena, we have to somehow a priori analyze mental truths into physical truths. While talk of *a priori analysis* has

largely disappeared in recent times, the core idea that reductive explanation/determination requires inferrability continues to thrive. Its most prominent proponents are Jackson (1998, 2001, 2003a, 2006, 2010), Chalmers (1996, 2001, 2009, 2012) and (a later version of) Lewis (1994). Jackson, for one, tries to capture his idea with claims like:

...if we have something which is nothing over and above an aggregation of elements of such and such kinds, the properties of that aggregation had better be a priori determined by the way the such and such elements are aggregated and the nature of what is aggregated.(2006, 240)

Chalmers argues for much the same view. In particular, in recent work (2012) he argues that ideal agents can infer (or as he puts it, “scrut”) *every truth whatsoever* from the grounding conjunction of the fundamental microphysical truth and the fundamental phenomenal truth (plus a couple of additional clauses, see Chalmers 2012). Chalmers is well aware that his argument not only supports something close to inferentialism, but also has important implications for the epistemology of metaphysics. For he says that his argument supports “Conceptually guided global metaphysics”, and he describes his approach with passages like:

*Conceptually guided global metaphysics*: using conceptual relations as a guide to metaphysical relations among truths about reality.(2012, pp. 441)

Metaphysical relations among truths about reality do not float free of conceptual relations, but are heavily constrained by them.(2012, pp. 442)

Chalmers is also the first to recognize that inferentialism offers a way to investigate the fundamental (2012, esp. sec. 8.2 & appendix 16). As he puts it,

The relation between the fundamental and the nonfundamental requires careful philosophical analysis. The analysis of this relation can play a significant role in constraining which truths are fundamental in turn....I concentrate on the role that scrutability [read: inferrability] and related notions can play in this project.(pp. 442)

To this end, he specifically argues for “The Fundamental Scrutability Thesis”, which is the claim that we (or at least idealized versions of us) can infer all truths from the metaphysically

fundamental truth.<sup>23</sup> Chalmers therefore places a constraint on accounts of the fundamental. On his view, whatever the fundamental truths might be, it must be the case that we can infer *all* truths from them.

### **3) The Case for Inferentialism**

#### **3.1 A Positive Argument**

At first glance inferentialism can seem quite plausible. For one thing, there's a compelling *prima facie* argument in favour of this theory. This is because, to return to the opening quotation, grounding relations are typically taken to be “a distinctive kind of metaphysical explanation” (Fine, 2012). For if X wholly explains Y in grounding Y, then it seems that X must fully account for Y's presence and nature. Doesn't a candidate explanation that fails to be a full account also fail to explain something? Consequently, it looks to follow that if I grasp X, I thereby grasp all the information I need to understand why it is that Y obtains and has the nature it does. But when I have all this information, I (or at least idealized version of me) seem to be in a good position to infer Y. For it seems like I can then read off Y's occurrence and nature from the information I possess.

Chalmers and Jackson offer arguments like this in favour of their conclusion that reductive explanation requires inferrability (see: Chalmers & Jackson 2001; Jackson 2003, 2006; Chalmers 2012).

#### **3.2 Scarce Counterexamples**

Another point in favour of inferentialism is that it's hard to find counterexamples. In

---

<sup>23</sup> As he puts it, “The Fundamental Scrutability Thesis says that all truths are scrutable from the metaphysically fundamental truths”. (2012, pp. 404)

particular, inferentialism about grounding avoids the sort of counterexamples that doom the analogous position about necessitation.

**Inferentialism about necessitation (IAB):** X necessitates Y only if Y can be inferred from X.

While IAB can also seem plausible at first, it faces many obvious counterexamples. For instance: i) *frogs are slimy* necessitates  $2+2=4$ , but we cannot infer  $2+2=4$  from *frogs are slimy*; and ii) *some bachelors are married* necessitates *Asia is the largest continent*, but we cannot infer the latter from the former. Such examples spring from the fact that anything trivially necessitates a necessary truth, and necessary falsehoods trivially necessitate anything.

While such examples defeat IAB, they do not trouble inferentialism about grounding. This is because all such examples are instances of necessitation without grounding. For instance, even though *frogs are slimy* necessitates  $2+2=4$ , it's not the case that *frogs are slimy* grounds  $2+2=4$ . And while *some bachelors are married* necessitates *Asia is the largest continent*, it's not the case that the former grounds the latter. The root explanation here is that while there are what seem to be trivial instances of necessitation that clearly do not entail inferrability—and so rebut inferentialism about necessitation—there is no such thing as a trivial instance of grounding.

### 3.3 Controversial Counter-examples

Another point in favour of inferentialism is that whenever we find an example with the potential to conflict with inferentialism, we typically also find that this example faces marked philosophical resistance on multiple fronts. For instance, we have a counter-example to inferentialism if moral truths are grounded in but cannot be inferred from non-moral truths. We likewise have a counter-example if phenomenal truths are grounded in but cannot be inferred

from the physical truth. The problem is that both of these potential counter-examples face entrenched philosophical resistance. Take the case of moral truths as a model. In order to use moral truths in a counter-example to inferentialism, we need to show that: i) non-moral truths ground moral truths, and ii) we cannot infer moral truths from non-moral truths. But both of these claims face vigorous resistance—most of which is unconnected to the debate over inferentialism. On one hand, there's plenty of non-cognitivist (e.g., Blackburn 1993, 1998; Gibbard 2003) and error-theoretic (e.g., Mackie 1977; Joyce 2001) (and perhaps non-naturalist, e.g., Schafer-Landau, 2003) hostility to the idea that non-moral truths ground moral truths. And on the other hand, writers like Jackson (1998, 2003) argue that we *can* infer moral truths from appropriate collections of non-moral truths. We face the same worries in the case of phenomenal truths. On the one hand, some people (e.g., Chalmers 1996, 2002, 2009; Alter 2007, forthcoming; Adams 2013) argue that physical truths do not ground phenomenal truths. And on the other hand, writers like Lewis (1994), Jackson (2005, 2006, 2010) and Boutel (2013) argue that we (or at least idealized versions of us) can infer phenomenal truths from physical truths. The end result is that, in order to press these moral and phenomenal counter-examples to inferentialism, we have to wade into difficult and uncertain battles that have already been raging for a long time.

#### **4) The Case Against Inferentialism**

##### **4.1 Macrophysical Truths**

While many potential counter-examples to inferentialism sink in the mire of entrenched philosophical controversy, I at least one is on firmer footing: that of the metaphysical and



inferential relationships between the complete microphysical truth and macrophysical truths.

To be precise here, the complete microphysical truth, P, is the conjunction of all the fundamental microphysical truths of our world (Chalmers 2012; Chalmers & Jackson, 2001; Block and Stalnaker, 1999; Levine, 2010).<sup>24</sup> Fundamental microphysical truths are truths about the fundamental microphysical entities, their physical properties, and/or the laws that govern them, all phrased in the conceptual vocabulary of microphysics.

For our purposes, a macrophysical truth is just a physical truth about macroscopic objects, properties and/or laws. Macrophysical truths therefore come in a macrophysical conceptual vocabulary supplied by folk theory and the sciences; in particular, the special sciences. Example macrophysical truths include: *granite is an igneous rock*, *zebras form herds* and *oaks have scalloped leaves*.

Micro and macrophysical truths are important because (according to virtually everyone<sup>25</sup> who accepts the very notion of grounding) the complete microphysical truth, P, wholly grounds each macrophysical truth. This means that part of a counter-example to inferentialism is already secure and free of controversy. All we need to now argue is that we (even idealized versions of us) cannot infer macrophysical truths from P. If we establish this conclusion, then we falsify inferentialism.

For shorthand, it's easy to formalize this argument:

- 1) If inferentialism is true, then, by definition, X wholly grounds Y only if we (or at least idealized agents) can infer Y from X.
- 2) We cannot infer macrophysical truths from P.
- 3) Nonetheless, P wholly grounds every macrophysical truth.

---

<sup>24</sup> If there is no fundamental microphysical level because our world exhibits infinite metaphysical descent, then P includes every microphysical truth below some intuitively 'low' level.(Chalmers, 2012)

<sup>25</sup> Again, priority monists are the only potential dissidents.

-----

4) Therefore, inferentialism is false.

It follows that inferentialism is true only if we can infer macrophysical truths from P. That is, Inferentialism is true only if InferM is true.

**InferM:** We (at least idealized versions of us) can infer macrophysical truths from the complete microphysical truth, P, that grounds them.

So we now have an opportunity: if we can convincingly argue that we cannot infer macrophysical truths from P—viz., if we can argue against InferM—then we can also convincingly argue that inferentialism is false.

#### 4.2 Extant Arguments and Setup

The importance of micro-macro inferential relations has not gone unnoticed. Much of the extant literature on reductive explanation and inferrability focuses on the question of whether we can infer macrophysical truths from the microphysical truth that grounds them. On one side people like Block & Stalnaker (1999), Byrne (1999), Tye (2009) and Levine (2010) all argue that even idealized subjects cannot infer macrophysical truths from P. And on the other side, Lewis (1994), Jackson (1998, 2001, 2003, 2006, 2010), and Chalmers (1996, 2001, 2009, 2012) argue that idealized subjects can infer macrophysical truths from P.<sup>26</sup>

While I obviously agree with Block, Stalnaker, et al.'s conclusion, this paper is not a defence of their arguments. Instead I develop a new argument for the conclusion that even

---

<sup>26</sup> I'm well aware that Chalmers and Jackson sometimes replace P with PTI or PQTI—because they hold that, in the final analysis, it's PTI or PQTI that grounds all macrophysical truths, not P itself. These additions are: i) *I*, a dose of locating information that's supposed to enable Mary to infer indexical macrophysical truths, ii) *T*, a 'that's-all' or totality clause which states (roughly) that P is the complete microphysical truth and thereby ensures that PT necessitates even negative and universally quantified macrophysical truths, and iii) *Q*, a complete statement of the fundamental phenomenal truths. (Chalmers & Jackson 2001; Chalmers 2002, 2004, 2009, 2012). **Feel free to substitute PTI or PQTI for P at any point in this paper.** This added complication does not change my argument. Because of this fact, I omit *T*, *I*, and *Q* for the sake of simplicity.

though P wholly grounds macrophysical truths, we cannot infer macrophysical truths from P. I do this because I think that Block, Stalnaker, et al.'s arguments are problematic. This is because, I argue, these arguments are only effective against an interpretation of InferM that's not only less plausible than alternatives, but also an uncharitable interpretation of Chalmers, Jackson, and Lewis. Thus, Block, Stalnaker, et al.'s arguments are inadequate and off-target.

To effectively argue against InferM, we need to develop a new argument against more charitable and plausible interpretations. This paper is my attempt to do just that. Here's how I'll proceed. First I'll explain: i) one interpretation of InferM, and ii) Block, Stalnaker, et al.'s arguments it. In doing so, I'll make it clear why Block, Stalnaker, et al.'s arguments are effective against the interpretation in question. I'll then explain a family of more plausible interpretations of InferM, and detail why they seem to avoid Block, Stalnaker, et al.'s arguments. I'll also provide evidence that these new interpretations are more charitable readings of Lewis, Chalmers, and Jackson. Finally, I'll develop my own argument against these further interpretations of InferM.

As such, my overall case against InferM and inferentialism unfurls in stages. Each stage coincides with a different interpretation of InferM; viz., with a different account of what it is for us to be able to infer macrophysical truths from P. My aim is to argue that on each of these accounts, it turns out that we cannot infer macrophysical truths from P. That is, via a strategy of divide-and-conquer, I argue that InferM is false regardless of interpretation.

### **4.3 Interpretation #1**

The most literal interpretation of InferM makes it the claim that P provides us with all the information that we need to infer Ms. This entails that even if we were informed solely by P—

i.e., if P were all the information we knew, believed, and/or assumed—we would still be able to infer Ms. As such, on this interpretation, InferM should really be read as:

**InferM-1:** We (or at least idealized versions of us) can infer macrophysical truths from P *alone* (i.e., from just the information P provides).<sup>27</sup>

With this reading in mind, now let's look at Block, Stalnaker, et al.'s arguments. There are two lines of thought here. The first (implicit in Tye (2009), Byrne (1999), and Levine (2010)) keys on the fact that there's a gap between the conceptual vocabularies of P and macrophysical truths. This gap exists because while P, *qua* fundamental microphysical truth, only involves fundamental microphysical concepts, macrophysical truths involve macrophysical concepts. The argument then claims that this conceptual gap entails a corresponding inference gap: i.e., the conclusion is that we cannot infer macrophysical truths from P because these truths come in different conceptual vocabularies.

The second primary argument (which originates with Block & Stalnaker (1999), and has since been put forward by Tye, Byrne, Levine and others) also starts with the fact that there's a gap between the conceptual vocabularies of P and macrophysical truths. But this argument then contends that while we might be able to infer our way across this conceptual gap, we can only do so if we can explicitly analyze macrophysical concepts into microphysical concepts. The argument then simply contends that macrophysical concepts cannot be explicitly analyzed into microphysical concepts. So the basic idea here is that since such analyses are unavailable, we cannot infer macrophysical truths.

I think that these two arguments are effective against InferM-1. This is because I agree

---

<sup>27</sup> Remember, one can substitute PTI or PQTI for P at any point. Thus, one can read InferM-1 as “We can infer macrophysical truths from PQTI alone”. This is because the core idea here is that we can infer macrophysical truths from *just* the base that grounds them—without any additional information playing any type of role.

that the conceptual gap between P and macrophysical truths renders us unable to infer macrophysical truths from P *alone*.

To explain why—and thereby flesh out some reasoning that I think is important to both of the above arguments—let's start with the fact that in order to infer macrophysical truths (from any inference base) one must possess the requisite macrophysical concepts.<sup>28</sup> That is, to infer macrophysical truths, one must possess the concepts that are their conceptual constituents (and, so, that refer to macrophysical objects, kinds, properties, etc.). So, for example, *salt is crystalline when solid* is a macrophysical truth, and *salt* and *crystalline* are component macrophysical concepts. Thus, the present point is simply that one must possess *salt* and *crystalline* in order to infer *salt is crystalline when solid* from any sort of inference base.

Everyone in the debate accepts this point. Even Chalmers and Jackson admit as much.

For whenever they make claims like:

Importantly, no other empirical information is required to justify the inference from PQTI to M [i.e., to macrophysical truths]: PQTI contains all the information that is needed to know M. (2001, pp. 16)

They also say (in this case, immediately afterwards) that:

All the is required here [i.e., to infer macrophysical truths] is *possession of the concepts in M*, the information in PQTI, and rational reflection. (2001, pp. 16-17, italics mine)

The reason everyone accepts this point is that it seems clear one needs to possess macrophysical concepts in order to even understand and entertain macrophysical truths. And, surely, inferring a truth requires the ability to understand and entertain it.

Now return to the idea that there's a conceptual gap between P and macrophysical truths.

---

<sup>28</sup> This does not mean that, in order to infer Ms, one must possess macrophysical concepts in an expert way. For example, I do not argue that one can infer macrophysical truths only if one possesses macrophysical concepts to the point of being able to define such concepts, or to recognize objects that fall into the extension of such concepts. I just argue that, in order to infer Ms, one must possess macrophysical concepts in some way.

This gap exists because P is just the complete microphysical truth in the conceptual vocabulary of microphysics. This means that P only involves microphysical concepts that pick out features of fundamental microphysical reality. As such, macrophysical concepts like *salt*, *elephant*, *crystalline*, and *mountain*—i.e., the sort of concepts that feature in macrophysical truths—do not feature in and are not the topic of P.<sup>29</sup> Such concepts just are not part of the fundamental microphysical description of the world.

The conceptual gap between P and macrophysical truths is important because it entails that if we're informed solely by P—i.e., if P is all the information we possess—then we not only fail to know, believe, or assume anything about the meaning, use, and etiology of macrophysical concepts, but we also lack knowledge, beliefs, or assumptions that involve these concepts as constituents. In short, if P is our sole source of information, then we lack any and all information regarding macrophysical concepts.

So now we have to wonder: given the conceptual gap between P and macrophysical truths, can we even possess macrophysical concepts if we're informed solely by P? The answer seems to be “no”. After all, if a subject is unaware of not only the meaning, use, and origin of a macrophysical concept like *granite*, but also of any and all truths that involve this concept as constituent, then, plausibly, she lacks the sort of information that she needs to possess *granite* (both of Block, Stalnaker, et al.'s arguments seem to ably capture this thought). This point gets even stronger when we consider that—as Block, Stalnaker, et al.'s second argument suggests—*granite* cannot be analyzed into the microphysical concepts we draw from P. For this entails that

---

<sup>29</sup> The same is true with regards to I, T, and Q. They only involve indexical, microphysical, and fundamental phenomenal concepts. See fn. 9. Thus, my point holds whether we fix on P, PTI, or PQTI as our putative grounding and inference base.

we cannot use conceptual analysis to access P as a source of information regarding *granite*.

Conclusion: the conceptual gap between P and macrophysical truths ensures that we cannot possess macrophysical concepts when we're informed solely by P.

This conclusion entails that we cannot infer macrophysical truths from P alone. That is, the conceptual gap between P and macrophysical truths falsifies InferM-1.

**InferM-1:** We (or at least idealized versions of us) can infer macrophysical truths from P *alone* (i.e., from just the information P provides).

For since we cannot even possess macrophysical concepts when we're informed solely by P, we cannot thence infer macrophysical truths.

#### 4.4 Are We Done Already?

It's tempting to think that when we show that we cannot infer macrophysical truths from P *alone*, we thereby show that we cannot infer macrophysical truths from P. In other words, it's tempting to think that when we falsify InferM-1, we thereby falsify InferM:

**InferM:** We (or at least idealized versions of us) can infer macrophysical truths from the complete microphysical truth, P, that grounds them.

After all, by arguing that we cannot infer macrophysical truths from P alone, what we've shown is that we can only hope to infer macrophysical truths when we're informed by not only P, but also by some extra information that helps enable us to possess macrophysical concepts. And it's natural that if we need any information beyond P in order to infer macrophysical truths, then there's no sense in which we can infer such truths *from P*.

*But this is a mistake.* Even though we cannot infer macrophysical truths from P *alone*, there is another important sense in which we might still be able to infer macrophysical truths from P. That is, there is another important interpretation of InferM that might be true even

thought InferM-1 is false.

This interpretation start with the well-worn idea that there are two different ways a piece of information can help us perform an inference. First, this information can help enable us to possess whatever cognitive, conceptual, and/or rational capacities we require in order to perform the inference. I'll say that when information performs this function, it plays an *enabling role*; and, so, that this information *enables*. Second, a piece of information can serve as one of the premises, data points, and/or pieces of evidence that helps justify our inference. I'll say that when information performs this function, it plays a *justifying role*; and, so, that this information *justifies*.<sup>30</sup>

It's plausible that pieces of information can justify without enabling and enable without justifying (though information can also play both roles). In other words, the information that equips us with the conceptual and cognitive capacities we need to draw an inference is not always part of the premises that actually justify our inference; and vice-versa. To illustrate this possibility, consider a simple example:

Hawaii is a volcanic archipelago

-----

Therefore, either Hawaii is a volcanic archipelago, or Mauna Kea is the highest mountain in the Pacific Ocean.

This simple inference is just an instance of disjunction introduction, and is clearly a justified inference that most anyone can draw. What's more, to draw this inference, I need to possess concepts like *mountain* and *ocean*. I therefore need information which enables me to possess such concepts. Truths like *Mountains are geographic high points* and "*Ocean*" refers to large

---

<sup>30</sup> It's important to see that with my labels "enabling role" and "justifying role" I describe types of roles, not tokens. There are many different ways to play an enabling or justifying role.



*bodies of salt water* are the sorts of things that might fill this enabling role. However, even if I believe or assume such truths and they thereby enable my inference, it seems clear that they do not justify it. That is, it seems clear that they do not serve as my justifying premises. For the lone truth “Hawaii is a volcanic archipelago” is sufficient to justify my inference.<sup>31</sup>

Examples like this suggest a way that we might reconcile InferM with our result that we need information beyond P if we're to have any hope of inferring macrophysical truths. This is because all we've shown so far is that we need extra information to help *enable us* to possess macrophysical concepts. So even though we need this extra information in order to infer macrophysical truths, it might still be the case that P alone is sufficient to *justify* any inference of macrophysical truths. That is, it might still be the case that P alone provides all the justifying premises, evidence, etc., that we need to infer macrophysical truths.

This is an important point because it seems quite natural to interpret InferM as just claiming that P alone suffices to justify any inference of macrophysical truths. That is, it seems quite apt to read InferM as something like:

**InferM-2:** We (or at least idealized versions of us) can infer macrophysical truths when: i) P alone supplies our justifying premises—i.e., the information available to play a justifying role in our inferences, and ii) we also have some extra information that enables us to possess the requisite macrophysical concepts.

Crucially, InferM-2 is compatible with the falsity of InferM-1. For even if we cannot infer macrophysical truths from P alone, it might still be the case that P provides us with all of the justifying premises that we need to infer such truths.

What's more, there's reason to think that InferM-2 might be the more charitable

---

<sup>31</sup> However (to illustrate the reverse point), this truth is not sufficient to enable my inference. This is because: i) this inference demands that we possess concepts like *mountain* and *highest*, and ii) a grasp of “Hawaii is a volcanic archipelago” does not itself enable us to possess such concepts.

interpretation of Chalmers and Jackson's arguments. For, remember, they say things like:

Importantly, no other empirical information is information is required to **justify** the inference from *PQTI* to *M*....All that is required here **is possession of the concepts in M**, the information in PQTI, and rational reflection. (2001, pp. 16-17, highlights mine)

Quotes like this one give us two options: interpret Chalmers and Jackson as i) proponents of something like InferM-2, or ii) as proponents of the idea that we can possess macrophysical concepts when we're informed by P alone. Since the latter idea is false, the first interpretation seems much more charitable—not to mention that it also respects Chalmers and Jackson's talk of how PQTI suffices to *justify* our inferences of macrophysical truths. Thus, it seems best to conclude that Chalmers and Jackson intend to argue for something like InferM-2.

#### 4.5 A Problem with Extant Arguments

As I said before, several other writers attempt to argue that we cannot infer macrophysical truths from P. The prime examples here are Block & Stalnaker (1999), Byrne (1999), Tye (2009) and Levine (2010). Two main arguments emerge from these attempts. The first keys on the fact that there's a gap between the conceptual vocabulary of P and macrophysical truths. The argument then claims that this conceptual gap entails that we cannot infer macrophysical truths from P.

The second primary argument also starts with the claim that there a gap between the conceptual vocabulary of P and macrophysical truths. But this argument states that while we might be able to infer our way across this gap, we can only do so if we can explicitly analyze macrophysical concepts into microphysical concepts. The argument then contends that since such analyses are unavailable, we cannot infer macrophysical truths from P.<sup>32</sup>

---

<sup>32</sup> This arguments originates with Block & Stalnaker (1999), and has since been put forward by Tye, Byrne, Levine and others.

As we've seen, both of these arguments are effective against InferM-1.

**InferM-1:** We (or at least idealized versions of us) can infer macrophysical truths from *P alone* (i.e., from just the information *P* provides).

This is because both arguments capture powerful reasons for why we cannot even possess macrophysical concepts when we're informed by *P* alone.

But, now, are these arguments also effective against InferM-2?

**InferM-2:** We (or at least idealized versions of us) can infer macrophysical truths when:  
i) *P* alone supplies our justifying premises—i.e., the information available to play a justifying role in our inferences, and ii) we also have some extra information that enables us to possess the requisite macrophysical concepts.

I suggest that, no, Block, Stalnaker, et al.'s arguments are not effective here. This is because, for one thing, these arguments cannot work against InferM-2 in the same way that they work against InferM-1: i.e., by pointing out why it is that we cannot possess macrophysical concepts if we're informed by *P* alone. For InferM-2 is perfectly compatible with the conclusion that we cannot possess macrophysical concepts when we're informed by *P* alone. In fact, InferM-2 specifically accommodates this conclusion, as it's a claim about what we can infer when we have all the enabling information we need to possess macrophysical concepts.

I also do not see any other way for Block, Stalnaker, et al.'s arguments to seriously challenge InferM-2. Consider, for instance, their first argument. It fixes on the fact that there's a conceptual gap between macrophysical truths and *P*. But with InferM-2, we're now asking whether we can infer macrophysical truths when we're informed by both *P* and some extra enabling information. So the gap between macrophysical truths and *P* does not amount to a conceptual gap between macrophysical truths and the inference base now at issue. As such, the conceptual gap between macrophysical truths and *P* seems irrelevant to the fate of InferM-2.

We also cannot easily change and update Block, Stalnaker, et al.'s first argument to work against InferM-2. For in order to do so, there has to be a conceptual gap between macrophysical truths and the union of P and enabling information. But there likely is no such gap, for any cache of information that enables us to possess macrophysical concepts is likely to include information about and/or involving macrophysical concepts.

Now consider their second argument. It fixes on the claim that macrophysical concepts cannot be analyzed into the microphysical concepts we draw from P. This is supposed to show that we cannot bridge the conceptual gap between P and macrophysical truths by using conceptual analysis as a way to access P as a source of information regarding M-concepts. But the failure of macro-->micro conceptual analysis seems unimportant now that we're addressing InferM-2. For as I just pointed out, there's likely no conceptual gap between macrophysical truths and the union of P and whatever extra information enables us to possess macrophysical concepts. Without such a gap, we don't need the bridge that might be supplied by conceptual analysis. That is, there's little reason to think that when we're supplied with both P and extra enabling information—and thereby have a firm grasp of macrophysical concepts—we still need to analyze macrophysical concepts into the microphysical concepts we draw from P in order to infer macrophysical truths. As such, even if there are no successful macro-->micro analyses, this does not seem to be a significant problem for InferM-2.

In conclusion, while Block, Stalnaker, et al.'s arguments work against InferM-1, neither seems very effective against InferM-2. In other words, neither effectively show that we cannot infer macrophysical truths when P alone is our source of justifying premises, and we also grasp some extra information that enables us to possess macrophysical concepts. To press our case, we

need a new argument.

## 5) Interpretation # 2

It's now clear that to convincingly argue that we cannot infer macrophysical truths from P, we need to argue that we cannot do so even when we're informed by both P and some extra information that enables us to possess macrophysical concepts. That is, we need to somehow refute our second and more charitable interpretation of InferM:

**InferM-2:** We (or at least idealized versions of us) can infer macrophysical truths when:  
i) P alone supplies our justifying premises—i.e., the information available to play a justifying role in our inferences, and ii) we also have some extra information that enables us to possess the requisite macrophysical concepts.

To do this we need to argue that possession of macrophysical concepts does not equip us to infer macrophysical truths from justifying information supplied by P. For if we show that this is true, we show that even if we have all the enabling information we need to wield macrophysical concepts, we cannot infer macrophysical truths from justifying information supplied by P. It then follows that not only are we unable to infer macrophysical truths from P alone (i.e., not only does P fail to *both* enable and justify such inferences), but P alone does not even justify inferences of macrophysical truths. That is, it follows that P alone does not provide us with information, data-points, and/or evidence that suffice to justify inferences of macrophysical truths. And this surely means that we (not even idealized versions of us) cannot infer macrophysical truths from P.

### 5.1 Many Theories of Concept Possession

What complicates this strategy is the fact that there are several theories of what it is to possess a

concept. This means that there's several competing accounts of what our extra information must do to enable us to possess macrophysical concepts. Since I don't want to adjudicate between various theories, I therefore need a way to evaluate InferM-2 across several different theories of concept possession.

Here's how I'll proceed: first I sort theories of concept possession into two categories: what I call i) undemanding theories, and ii) demanding theories. This strategy allows me to consider many theories of concept possession as one. To draw this distinction I sketch a scenario where my protagonist is only informed by P plus a very scant collection of extra information about some macrophysical concepts. So while my subject has some enabling information beyond P, she only has a little bit. A theory of concept possession is *undemanding* just in case it entails that my subject possesses the macrophysical concepts in question. Accordingly, a theory is *demanding* just in case it entails that my subject does not possess the concept at issue. I then argue two things: i) that many theories of concept possession are undemanding, and ii) that on any undemanding theory, InferM-2 is false. Finally, to finish my campaign of divide-and-conquer, I consider demanding theories in turn, and argue that each also falsifies InferM-2. Conclusion: InferM-2 is false on every theory of concept possession.

## 5.2 Undemanding Theories

Consider the case of Judy. She's a cognitively idealized subject who participates in a community of people who employ and are often well informed about the macrophysical concepts *elephant* and *tusk*.<sup>33</sup> But despite this social contact, she is deeply uninformed about these concepts. In fact, P is Judy's predominant source of information. Beyond P, the only truths she

---

<sup>33</sup> When I say Judy is idealized, I mean that she has an ideal intellect. I do not mean that she has an idealized knowledge-set—i.e., that she's near omniscient.

knows, believes and/or assumes are: i) logical truths like *All elephants are elephants*, and *If x is a tusk, then x is a tusk*. and ii) the bland generalities *Elephants are a type of animal*, and *Tusks are a type of object*. To be clear, she does not have any other beliefs or assumptions that involve *elephant*, *tusk* or any other macrophysical concepts as conceptual constituents. Nor does she have any higher-order beliefs or assumptions about the meaning, use, etiology etc., of macrophysical concepts.<sup>34</sup>

Does Judy possess the macrophysical concepts *tusk* and *elephant*? Many popular theories of concept possession entail that she does. In other words, many popular theories are what I call, *undemanding theories of concept possession*. One such is the “Cartesian” (Fodor, 2004) account of concept possession. This is theory that one possesses a concept C just in case one can exercise C in thought (see, e.g., Fodor 1998, 2004; Ball 2009). Or, in the words of Fodor:

To have the concept DOG is to be able to think about dogs as such; and conversely, to be able to think about dogs as such is to have the concept DOG. (2004, pp. 3)

Since it's clear that Judy can exercise *elephant* and *tusk* in thought—after all, she must have done so to form beliefs/assumptions that involve these concepts—it's also clear that, on Cartesianism, she possesses these macrophysical concepts.

Another class of undemanding theories are those which state that possessing a concept is a matter of having knowledge, beliefs, and/or assumptions which involve that concept as a constituent (e.g., Bealer, 1998, Peacocke 2008). For since Judy believes or assumes *Elephants are a type of animal*, and *Tusks are a type of object*, such theories entail that she possesses *elephant* and *tusk*.

---

<sup>34</sup> This means that she lacks belief and assumptions involving second-order concepts that refer to *elephant* and *tusk*.

Externalist theories of concept possession likely also entail that Judy possesses *elephant* and *tusk* (see, e.g., Putnam 1970, 1975; Burge 1979, 1986; Williamson 2006).<sup>35</sup> This is because such theories state that even if one is largely ignorant about a concept, one can nonetheless possess it if one participates in a linguistic community that uses this concept and, collectively, knows a lot about it. As such, on externalism, the fact that Judy participates in a community of *elephant* and *tusk* users strongly suggests that she also possesses these concepts.

As we can now see, many popular theories of concept possession allow Judy to possess the macrophysical concepts *elephant* and *tusk*. So on all of these theories, Judy both: i) has P as a source of premises—i.e., of justifying information—and, ii) possesses the macrophysical concepts *elephant* and *tusk* (by virtue of having a limited set of information beyond P). With this in mind, look again at InferM-2.

**InferM-2:** We (or at least idealized versions of us) can infer macrophysical truths when: i) P alone supplies our justifying premises—i.e., the information available to play a justifying role in our inferences, and ii) we also have some extra information that enables us to possess the requisite macrophysical concepts

It follows on any undemanding theory of concept possession, if InferM-2 is true, then Judy (remember, she's a cognitively idealized subject) can infer some macrophysical truths that involve *elephant* and *tusk*. For instance, InferM-2 entails that she can infer macrophysical truths like, *Elephants sometimes have tusks*.<sup>36</sup>

But is this plausible? In her present position, can Judy really infer such macrophysical truths? I submit that, no, she cannot. For she lacks the information she requires to perform such

---

<sup>35</sup> I say “likely” because it's consistent to be an externalist about concept possession and yet argue that Judy does not possess the concepts in question. However, I know of no extant externalist theory which entails that Judy lacks these concepts.

<sup>36</sup> Here I assume that her possess of P provides her with the concepts *sometimes* and *have*. If you disagree, just provide with the information she needs to possess these concepts.



inferences. Remember, while she may possess *elephant* and *tusk*, she has very little information regarding these and other macrophysical concepts. Beyond P, all she believes or assumes are trivialities and two bland generalities (viz., *Elephants are a type of animal*, and *Tusks are a type of object*). With only this scant information, she cannot recognize or understand the connections: i) between *elephant* and *tusk*, ii) between these macrophysical concepts and relevant microphysical concepts, and iii) between macrophysical truths involving these concepts and relevant microphysical truths. And without a sense of these connections, she does not have justification to infer truths like *elephants sometimes have tusks* from premises supplied by P.

The result is that on undemanding theories of concept possession like Cartesianism and various brand of externalism, Judy's case is a counter-example to InferM-2. For here we have a person who has P as her source of justifying information, possesses macrophysical concepts (because of her enabling information), and yet still cannot infer macrophysical truths.

Conclusion: on an undemanding theory of concept possession, InferM-2 is false.

### **5.3 An Important Qualification**

To be clear: with the above argument, I do not completely reject the idea that we might be able to infer macrophysical truths if, in addition to P, we have extra information about macrophysical concepts like *elephant* and *tusk*. All I argue (so far) is that if this extra information does nothing more than enable us to possess macrophysical concepts in the manner specified by undemanding theories of concept possession, then it does not help us infer macrophysical truths. For as Judy's case shows, this degree of macrophysical concept possession does not equip us to infer macrophysical truths from P. As such, there is still the open possibility that information beyond P can help us infer macrophysical truths from P when it enables us to

possess macrophysical concepts in a way that goes beyond undemanding macrophysical concept possession. I explore this possibility in the next few sections.

#### 5.4 A More Demanding Theory: Recognitional Concept Possession

So far I've argued that on an undemanding theory of concept possession—viz., one which entails that Judy possesses *elephant* and *tusk*—InferM-2 is false. But some theories of concept possession are more demanding. That is, there are some important theories which entail that Judy does not possess *elephant* and *tusk*. So now we have to consider the possibility that we can infer macrophysical truths when, in addition to P, we have extra information that enables us to possess macrophysical concepts in the manner specified by these more demanding theories.

This discussion is manageable because the literature only contains two types of demanding theories of concept possession. I'll now consider them in turn.

First we have *recognitional theories* of concept possession. These theories state that a person possesses a macrophysical concept to the extent that she has the ability to *accurately recognize* things that fall into the extension of this concept, and thereby distinguish them from things that don't (e.g., Evans 1982; Peacocke 1992). So on this picture, a person possesses an M-concept like *grizzly bear* if she's able to recognize grizzly bears as such, and can thereby sort grizzly bears from non grizzly bears.

Recognitional accounts of concept possession look to be demanding in my sense because Judy, with her scant information, probably does not have the ability to recognize elephants and tusks as such. Thus, Judy's scant information does not allow her to possess *elephant*, *tusk*, or any other macrophysical concept in a recognitional manner.

If we understand concept possession as being tied to recognitional abilities, then we

interpret InferM-2 as something like:

**InferM-2\*:** We (or at least idealized versions of us) can infer macrophysical truths when: i) P alone supplies our justifying premises—i.e., the information available to play a justifying role in our inferences, and ii) we also possess some extra information, E, that enables us to possess macrophysical concepts such that we're disposed to *recognize things that fall into the extension* of each macrophysical concept.

In doing so, we get an interpretation that's compatible with my example of Judy.

But is this new reading of InferM plausible? No. Even though Judy's case is not a counter-example to our new interpretation, it's still deeply problematic. The new problem is that there's no reason to think that recognitional possession of macrophysical concepts helps us to infer macrophysical truths from P. This is because there does not seem to be any close connection between the ability to recognize macrophysical objects, properties, etc., and the ability to infer macrophysical truths from microphysical truths. For while recognitional possession does come with the ability to learn all sorts of macrophysical truths by direct observation, this does not suggest that thus possessing macrophysical concepts helps us to infer macrophysical truths.

### 5.5 A More Demanding Theory: Inferential Concept Possession

The literature contains one more, very influential, suggestion about what it is to possess a concept. This is:

**Inferential Possession:** A person possesses a M-concept to the extent that she's disposed to draw and/or accept certain inferences that follow because of this concept's content (see, e.g., Brandom 2000, 2008; Peacocke 1992, 2008).<sup>37</sup>

This theory keys on the idea that while some inferences have a logical form which ensures that they follow whatever the content of the concepts involved, other inferences only follow because

---

<sup>37</sup> I use “follow” informally here. An inference that follows is just a good inference. So on my usage, an abductive, inductive, transcendental, etc., inference can follow even though it's not logically valid.

of the particular concepts and contents involved. For example, while logical form validates inferences from *Doug owns a dog and a cat* to *Doug owns a dog*, the same is not true for the inference from *Doug owns a poodle* to *Doug owns a dog*. In this case it's the contents of the macrophysical concepts *poodle* and *dog* which dictate that inferences from *Doug owns a dog* to *Doug owns a poodle* follow. On inferentialism, this is the sort of inference that matters for concept possession. In fact, an inferential theory might even say that if one is disposed to draw the inference from *Doug owns a poodle* to *Doug owns a dog*, then one possesses *poodle* and *dog* to some extent.

Inferential theories are demanding in my sense because Judy, with her scant information about *elephant* and *tusk*, likely is not disposed to draw inferences that follow because of the content of these concepts. Thus, Judy's scant information likely prevents her from inferentially possessing *elephant* and *tusk*.

If we understand concept possession as being tied to inferential dispositions, then we end up with yet another interpretation of InferM2. Something like:

**InferM-2\*\*:** We (or at least idealized versions of us) can infer macrophysical truths when: i) P alone supplies our justifying premises—i.e., the information available to play a justifying role in our inferences, and ii) we also have some extra information, E, that enables us to possess macrophysical concepts such that we're disposed to draw certain inferences that follow because of the contents of these concepts.

While this new reading escapes my previous arguments, it's nonetheless false. To see why, first note that our new interpretation is only plausible to the extent it's plausible that the contents of macrophysical concepts dictate that inferences from P to macrophysical truths follow. This is because if the contents of macrophysical concepts do not dictate that such inferences follow, then an inferential grasp of macrophysical concepts does not come with dispositions to

infer macrophysical truths from P.<sup>38</sup> And if an inferential grasp of macrophysical concepts does not come with such dispositions, then it's hard to see how an inferential grasp helps us infer macrophysical truths from P. Consequently, the plausibility of InferM-2\*\* is determined by the plausibility of the hypothesis that the contents of M-concepts dictate that inferences from P to macrophysical truths follow.

However, the claim that the contents of macrophysical concepts dictate that inferences from P to macrophysical truths follow is tantamount to the claim that we (or at least idealized versions of us) can infer macrophysical truths from P. For inferences that follow are inferences that we can justifiably draw if we're smart and energetic enough.

This means that the plausibility of InferM-2\*\* is effectively determined by the plausibility of the hypothesis that we (or at least idealized versions of us) can infer macrophysical truths from P. But, notice, this just means that the plausibility of InferM-2\*\* is determined by the plausibility of InferM.

**InferM:** We (or at least idealized versions of us) can infer macrophysical truths from the complete microphysical truth, P, that grounds them.

This result reveals that InferM-2\*\* and, thus, the notion of inferential concept possession, do not advance the dialectic whatsoever. That is, the inferential account of concept possession does not generate any new reason to think that, when we thus possess macrophysical concepts, we can infer macrophysical truths from P. For in order to figure out whether InferM-2\*\* is plausible, we need to first answer our original question of whether InferM is true. So we just end up back where we started.

---

<sup>38</sup> Of course, it's consistent for us to still have these dispositions, but the point is that an inferential grasp of macrophysical concepts is not going to provide us with them.

What follows from all this is we can assess InferM-2\*\* by simply assessing InferM in general. For InferM-2\*\* does not stand on its own as an interpretation of InferM. Instead, InferM-2\*\* derives any plausibility it has from the plausibility of InferM; and thus from the plausibility of other interpretations of this hypothesis. This entails that InferM-2\*\* is vulnerable to any argument we level at other interpretations of InferM. So if we show that every other natural interpretation of InferM is deeply implausible, then we also show that InferM-2\*\* is deeply implausible.

At this point in the paper, I can adopt just such a strategy against InferM-2\*\*. For I've now argued against not only:

**InferM-1:** We (or at least idealized versions of us) can infer macrophysical truths from P *alone* (i.e., from just the information P provides).

But also against several other versions of InferM-2:

**InferM-2:** We (or at least idealized versions of us) can infer macrophysical truths when:  
i) P alone supplies our justifying premises—i.e., the information available to play a justifying role in our inferences, and ii) we also have some extra information that enables us to possess the requisite macrophysical concepts.

As such, I think I've now argued against every other natural interpretation of InferM aside from InferM-2\*\*. I further submit that since my prior arguments succeed against these other interpretations, they also succeed against InferM-2\*\*. Conclusion: an inferential grasp of macrophysical concepts does *not* equip us to infer macrophysical truths from P.

## **5.6 Conclusion: Concept Possession Does not Help**

At this point I've now covered every extant type of theory of concept possession. And in each instance the result is the same: it's not the case that we can infer macrophysical truths when we're informed by both P and some extra information whose role is to enable us to possess

macrophysical concepts. In short, every theory of concept possession affirms that a possess of macrophysical concepts does not equip us to infer macrophysical truths from premises supplied by P. This means that, no matter what theory of concept possession we favour, InferM-2 is false.

**InferM-2:** We (or at least idealized versions of us) can infer macrophysical truths when: i) P is our sole source of premises—i.e., of justifying information, and ii) we also have some extra information, E, that enables us to possess the requisite macrophysical concepts. Importantly, E is only needed to play this enabling role.

## 5.7 Bringing it All Together

I started out by arguing that we cannot infer macrophysical truths from just the information provided by P. This is because P alone does not provide us with the *enabling information* that we need to possess macrophysical concepts. However, I then pointed out that this conclusion does not yet show that we cannot infer macrophysical truths from P. This is because even if we cannot infer such truths from P *alone*, it might still be the case that we can infer macrophysical truths when P is our source of justifying premises, and we also have some extra information that enables us to possess macrophysical concepts. In other words, it might still be the case that P alone suffices to *justify* our inferences of macrophysical truths. And I said before, if P alone suffices to justify our inference of macrophysical truths, then it's reasonable to say that we can infer such truths from P.

But what I've now just argued is that even when we have extra enabling information, and therefore possess and wield macrophysical concepts, we cannot infer macrophysical truths from premises supplied only by P. That is, P alone does not provide us with information, data-points, and/or evidence that suffice to justify inferences of macrophysical truths. So not only are we unable to infer macrophysical truths from P alone (i.e., not only does P fail to both enable and

justify such inferences), but P alone does not even justify inferences of macrophysical truths. This means that to have any hope of inferring macrophysical truths, we need extra information beyond P to help *justify* our inferences. Conclusion: we (even idealized versions of us) cannot infer macrophysical truths from P. InferM is false.

**InferM:** We (or at least idealized versions of us) can infer macrophysical truths from the complete microphysical truth, P, that grounds them.

The falsity of InferM is a fatal blow to inferentialism.

**Inferentialism:** X grounds Y only if Y can be inferred from X.

For now we see that while P, the complete microphysical truth, wholly grounds macrophysical truths, it's not the case that macrophysical truths can be inferred from P. Thus, we now have an important and damning counterexample to inferentialism; an example which shows that even though there's a *prima facie* compelling case for inferentialism this theory is false. Inference is not a guide to ground, and we cannot learn much about the structure or basis of reality by investigating the limits of inference.

## 6) An Alternative Thesis?

### 6.1 First Try

One thing an objector might argue is that while my argument works against inferentialism, we can formulate a related but alternative thesis in a way that accommodates my reasoning. And, importantly, this objector argues that with this alternative, we still get a theory that has important implications for the epistemology of grounding and the methodology of grounding-centric metaphysics.



To formulate a promising alternative to inferentialism, we need to accommodate my conclusion that, in order to infer macrophysical truths, we not only need extra information beyond P, but we also need this extra information to help justify our inferences. In short, we must accommodate the result that we need not only enabling but also *justifying information* beyond P. As such, we need to reformulate inferentialism as something like:

**Justify-1:** X wholly grounds Y only if Y can be inferred from XJ (where J is some set of additional truths that play both an enabling *and justifying role*).

The good thing about Justify-1 is that it's compatible with the grounding and inferential relations between microphysical and macrophysical truths. For I'm certain there's some J such that we can infer macrophysical truths from the conjunction of P and this J.

However, the problem with Justify-1 is that it's universally true. That is, no matter what X and Y are, and no matter the metaphysical relations between them—or lack thereof—there is some J such that we can infer Y from XJ. For instance, suppose that X is *Granite is an igneous rock* and Y is *Salt is crystalline when solid*. In this case, X does not ground Y. It's also evident that we (even when idealized) cannot infer this Y from this X. But now suppose that we believe a mass of extra information beyond *granite is an igneous rock* that enables my possess of the concepts *granite*, *igneous*, and *rock*, and is available to justify our inferences. Further suppose that our information includes:

- 1) *Granite is crystalline when solid because it's a mineral substance, and all mineral substances are crystalline in solid form.*
- 2) *Salt is a mineral substance.*

In this circumstance, we can infer *salt is crystalline when solid* from (1)-(2). Consequently, even though we cannot infer *salt is crystalline when solid* from *granite is an igneous rock*, we can

infer the former from the conjunction of (1)-(2) and *granite is an igneous rock*. This means that even though *granite is an igneous rock* does not ground *salt is crystalline when solid*, these truths satisfy the necessary condition on grounding imposed by Justify-1.

The fact that the above example satisfies Justify-1 is not itself the problem. This example is not supposed to be a counterexample, and it does not falsify Justify-1. No, the problem is that the sort of example on offer can be repeated with virtually any two truths. No matter what truths we choose, the consequent of Justify-1 is true. This entails that Justify-1 is not a philosophically interesting thesis. It doesn't help us in the epistemology of grounding, and it doesn't guide the practice of grounding-centric metaphysics. For since the consequent of Justify-1 is true for any X and Y, this thesis does not provide us with a useful test for whether X grounds Y. We get a positive test-result every time, no matter which X and Y we test, and regardless of the metaphysical relations between them. Similarly, since the consequent of Justify-1 is true for any X and Y, the finding that Y can be inferred from X no longer serves as any evidence that X grounds Y. In sum, though (or, really, because) Justify-1 is universally true, its truth is bought at the price of its philosophical importance.

## 6.2 Second Try

While Justify-1 is philosophically uninteresting, there is another nearby alternative to inferentialism that someone might propose. This is:

**Justify-2:** X wholly grounds Y only if there is some J (some truth or set of truths beyond X that play both an enabling and justifying role) such that Y can be inferred from XJ, *but not from J alone*.

Justify-2 is true. This is because if X grounds Y, then the claims *X necessitates Y*, *If X were true, then Y would be true*, and *If X, then Y* are true, and it's clear that: i) while we can infer Y from the

conjunction of X and one (or more) of these truths, ii) we cannot infer Y from one (or more) of these truths alone.<sup>39</sup> That is, if X grounds Y, then there are truths that, as J, satisfy the consequent of Justify-2.

Justify-2 is also of use in the epistemology of grounding and the methodology of grounding-centric metaphysics. This is because it offers us an interesting way to test whether X grounds Y: viz., we're to do this by figuring out whether *X necessitates Y*, *If X were true, then Y would be true*, and *If X, then Y* are true. If they are, then the relationship between X and Y passes Justify-2's necessary condition on grounding.

What's more, in offering this new test for grounding, Justify-2 does not succumb to the problem that sunk Justify-1: as the consequent of Justify-2 is not universally true. It's just not the case that for any given X and Y, there is some J such that we can infer Y from XJ but not from J alone. For example, while it's certainly true that we can infer *salt is crystalline when solid* from the conjunction of (1)-(2) and *granite is an igneous rock*, we can also infer *salt is crystalline when solid* from (1)-(2) alone. And as far as I can tell, there is no substitute for (1)-(2) which is such that we can infer *salt is crystalline when solid* from the conjunction of this substitute and *granite is an igneous rock*, but not from this substitute alone. Hence, *granite is an igneous rock* and *salt is crystalline when solid* do not satisfy the necessary condition on grounding that Justify-1 imposes. In light of such examples, we cannot wave Justify-2 aside on the grounds that its universal truth renders it philosophically unimportant.

So, have we finally found a thesis that's true *and* philosophically important? One that guides us in the epistemology of grounding and the methodology of grounding-centric

---

<sup>39</sup> Interestingly, *x grounds y* does not work here. For we can infer Y from this premise alone. Unlike statements of material implication or necessitation, statements of grounding assert the existence of both ground and grounded.

metaphysics? Yes, I think we have. But while this is a significant result, it's important to see that we're now left with a thesis that's very different from inferentialism. For Justify-2 does not focus on the issue of whether inferability is a guide to ground—viz., on the issue of whether the inferential relations between X and Y (or lack thereof) are a guide to underlying metaphysical relations. This is because Justify-2 reduces to a thesis like:

X grounds Y only if X necessitates Y, X materially implies Y, and X counterfactually implies Y.

That is, Justify-2 focuses on the issue of whether the modal and logical relations between X and Y (or lack thereof) are a guide to the metaphysical relations between them.

Of course, if it were true that we could effectively test the truth of *X necessitates Y*, *If X were true, then Y would be true*, and *If X, then Y* by checking whether we can infer Y from X, then Justify-2 could still be seen as a type of inferentialism. But this is just not so. Inferability is not a good test for necessitation, material implication, or counterfactual implication. For example, the three following claims are true: i) *“Egypt is a source of sodium carbonate decahydrate” necessitates “Egypt is a source of natron”*, ii) *If Egypt were a source of sodium carbonate decahydrate, then Egypt would be a source of natron*, and iii) *If Egypt is a source of sodium carbonate decahydrate, then Egypt is a source of natron*. Yet we cannot infer *Egypt is a source of natron* from just *Egypt is a source of sodium carbonate decahydrate*. We also need the information that natron is sodium carbonate decahydrate. So in this case, the inferential relationship—or, really, lack thereof—between our two truths is not a guide to the modal and logical relations between them.

What follows is that while Justify-2 is true and philosophically important, when we arrive

at this thesis, we effectively abandon inferentialism. For when we settle on Justify-2, we settle on a thesis which does not specify that inference is a guide to ground. Instead, we settle on a thesis which tells us something that we already knew: that modal and logical relations are a guide to ground. So, are there any viable alternatives to Inferentialism? Yes there are. But we only find such alternatives when we effectively abandon the core inferentialist idea that the inferential relations between two truths are an important guide to their metaphysical relations.

### Chapter 3: Mary's Macrophysical Ignorance

The thought example behind the knowledge argument is familiar to most. Our protagonist is Mary, a super-scientist with an idealized intellect. Mary has lived her whole life in a black and white room where she's received intensive instruction about the physical world by way of books, radio, telephone, and black and white television. We're asked to imagine that, through her study, Mary acquires vast knowledge of physical truths. However, Mary has never experienced shades other than white, black and grey. It's therefore intuitive that while trapped in her room, Mary is ignorant of what it's like to experience colours such as red, green, etc. That is, it seems that despite her intellect and education, Mary can only learn such phenomenal truths by escaping her room and experiencing colour.

The particulars of Mary's situation are familiar because they've been standardized for a long time. Three decades after Mary's appearance, everyone who writes about the knowledge argument still engages with her case. In contrast, the knowledge argument itself has changed with the years. It has been steadily updated to avoid objections and incorporate new philosophical ideas. In particular, in recent work Alter (2007, forthcoming) and Chalmers (2004, 2009) defend a sophisticated formulation of the knowledge argument that draws on two-dimensional theories of modality and content.<sup>40</sup> I call their formulation the *2-D Argument*.

In this paper I discuss and evaluate the 2-D argument. I first argue that it significantly advances the dialectic, as some popular responses to prior versions of the knowledge argument struggle to even challenge the 2-D argument. I then step into the breach and develop a new

---

<sup>40</sup> Jackson (1995, 2003) also talks about the knowledge argument in a way that's broadly in the spirit of the 2-D argument. However, he doesn't go into the same amount of detail as Alter and Chalmers.

objection to the 2-D argument. This objection follows a tack that's new to the dialectic on the knowledge argument. I argue that trapped-Mary's inability to learn certain phenomenal truths—however pronounced it turns out to be—is matched by her inability to learn ordinary, non-phenomenal macrophysical truths, e.g., *Granite is an igneous rock*, *Salt encrusts the surface of potato chips*, and *Bulldogs have jowls*. So, instead of trying to deny or explain away Mary's supposed ignorance of certain phenomenal truths, I argue that since she's equally ignorant of many macrophysical truths, any phenomenal ignorance on her part does not threaten physicalism.

In section 1, I explain the 2-D knowledge argument in relation to its predecessors. Section 2 argues that some notable objections to previous formulations of the knowledge argument fail to challenge the 2-D argument. Section 3 argues that the 2-D argument relies upon two theses:

**~Infer-Q:** In her room, Mary *cannot* infer phenomenal truths about experiences of colour (Qs) from the fundamental microphysical truth.

**Infer-M:** In her room, Mary *can* infer every non-phenomenal macrophysical truth (Ms) from the fundamental microphysical truth.

In short, I argue that the 2-D argument is built on the idea that there's an inferential asymmetry between phenomenal and non-phenomenal macrophysical truths. Section 4 then points out that extant objections to the 2-D argument ignore an opportunity: none of them challenge Infer-M. I step into the dialectical void in section 5, where I start by arguing against Infer-M. I argue that when Mary is still in her room, it's likely that she cannot master macrophysical concepts to the extent that's needed to infer all non-phenomenal macrophysical truths. I then argue that if we alter Mary's situation to make it plausible that she can infer non-phenomenal macrophysical

truths, it also becomes plausible that she can infer relevant phenomenal truths. In short, either Infer-M is false, or  $\sim$ Infer-Q is false. Either way, there is no inferential asymmetry between phenomenal and non-phenomenal macrophysical truths, and the 2-D knowledge argument is in trouble.

## **1) The Progression of the Knowledge Argument**

### **1.1 Two Formulations**

While the knowledge argument has changed with the years, prior to the 2-D argument it often took something like the following form (I call it the *Traditional Knowledge Argument*).

- 1) While still in her room, Mary gains complete physical knowledge.
- 2) Thus, while still in her room, Mary learns all of the physical facts.
- 3) Upon escape, Mary experiences colour for the first time and learns what it's like to experience colour—i.e., she gains new knowledge about the phenomenology of colour experiences.
- 4) Thus, upon escape, Mary learns facts about the phenomenology of colour experiences that are new to her.
- 5) Hence, there are non-physical facts about the phenomenology of colour experiences.
- 6) If physicalism is true, every fact about colour experiences is a physical fact.
- 7) Therefore, physicalism is false. (e.g., Jackson 1986; Nida-Rumelin 1998; Ball 2009)

This argument attacks physicalism by trying to show that phenomenal truths do not fall into the class of physical truths. This strategy puts emphasis on the issue of what Mary knows in her room.

In contrast, the 2-D argument puts emphasis on the issue of what Mary can infer while



trapped in her room. The argument springs from the fact that Mary's education is supposed to gift her with vast knowledge of physical truths. In particular, her education is surely designed to teach her the fundamental microphysical truth. It's then asked: what higher-level truths can Mary infer from the fundamental microphysical truth (i.e., P)? The 2-D argument starts with an answer to that question. Its first premise reads:

1) Mary cannot infer certain phenomenal truths, Q, (i.e., phenomenal truths about what it's like to experience colour) from P while still in her room.<sup>41</sup>

The argument then applies two-dimensional modal and semantic theory. For our purposes we can ignore the vast majority of this background. Only two important notions need to be defined.<sup>42</sup> The first is:

**Primary Necessity:** A truth T is primarily necessary if and only if T is true in all possible worlds considered as actual (Chalmers 2002, 2004, 2009).

From this definition we can then define a corollary notion:

**Primary Contingency:** A truth T is primarily contingent just in case T is false in some

<sup>41</sup> I'm well aware that proponents of the 2-D argument sometimes supplement P with two additional clauses that are designed to head off certain worries. These additional clauses are: i) *I*, a dose of locating information that's supposed to enable Mary to infer indexical macrophysical truths, and ii) *T*, a 'that's-all' clause which states (roughly) that P is the complete microphysical truth and thereby ensures that PT necessitates even negative and universally quantified macrophysical truths (Chalmers 2002, 2004, 2009). I omit these details because they're irrelevant to my argument. All the same, one can substitute PTI for P at any point in this paper. This added complication does not change the discussion.

<sup>42</sup> The basic idea of 2-D theory is that there are two ways of considering a possible world when we evaluate the truth-value of a claim in that world. First, we can consider a possible world *as actual*: that is, take it to be the actual world. Second, we can consider a possible world as counterfactual: thereby acknowledging that the actual world is a certain way and considering the possible world in question as merely an alternative way the actual world could have been. According to 2-D theory, a claim's truth value in a given world can vary according to how that world is considered (Chalmers 2002, 2006, 2009; Jackson 1998, 2004). One standard example involves the truth *water is H<sub>2</sub>O* and the infamous XYZ-world (i.e., a possible world that's just like our world except for the fact the clear, colourless, polar, etc., liquid which fills its oceans and rivers is XYZ, not H<sub>2</sub>O). When the XYZ world is considered as actual, *water is H<sub>2</sub>O* comes out false in that world. For if the XYZ world is the actual world, the concept *water* picks out not H<sub>2</sub>O but XYZ (and, indeed, does so rigidly in all possible worlds). However, when the XYZ world is considered as counterfactual, *water is H<sub>2</sub>O* is true there. For *water* rigidly refers in all possible worlds, and what it refers to is determined by its referent in the acknowledged actual world (in this instance, our world): which is H<sub>2</sub>O. This suggested variation leads some people to conclude that *water is H<sub>2</sub>O* true in all possible worlds when they're considered as counterfactual, but false in some possible worlds when they're considered as actual (Chalmers 2002, 2006, 2009; Jackson 1998; 2004).

possible world considered as actual (Chalmers 2004, 2009).

With these notions in hand, proponents of the 2-d knowledge argument offer the following thesis:

**2-D Thesis:** If T is not apriori, T is primarily contingent (Chalmers 2004, 2009).

The 2-D thesis is a crucial part of the 2-D argument. It makes a leap from claims about inference and a priority to conclusions about modality. This leap is important because the 2-D argument is designed to attack the foundational physical position that P (the fundamental microphysical truth) metaphysically necessitates all phenomenal truths. While many physicalists think that physicalism demands more than just physical necessitation of the phenomenal, almost everyone thinks that physicalism minimally entails such necessitation.

The 2-D knowledge argument is:

- 1) In her room, Mary cannot infer certain phenomenal truths, Q, from P (the fundamental microphysical truth).
- 2) Thus, Q cannot be inferred from P.
- 3) Hence,  $P \supset Q$  is not apriori.
- 4) If  $P \supset Q$  is not apriori,  $P \supset Q$  is primarily contingent. (From 2-D thesis)
- 5) If  $P \supset Q$  is primarily contingent,  $P \supset Q$  is metaphysically contingent.
- 6) If physicalism is true,  $P \supset Q$  is metaphysically necessary.
- 7) Therefore, physicalism is false. (Chalmers 2004, 2009; Alter, 2007, forthcoming)

The inference from (1) to (2) is based on the supposition that Mary is an idealized reasoner. The idea is that if she, *qua* ideal reasoner, cannot infer phenomenal truths from P, then it must be the case that phenomenal truths cannot be so inferred.

The inference from (2) to (3) is based on the fact that a  $P \supset Q$  material conditional is a

priori just in case i) Q is a priori, ii)  $\sim P$  is a priori, and/or iii) Q can be inferred from P. But in this instance, P is the microphysical *truth*, and phenomenal truths are not themselves a priori (if they were, Mary would be able to infer them in her room!). Thus, in effect, it turns out that  $P \supset Q$  is a priori just in case Q can be inferred from P.

Premise (5) is supported by various aspects of 2-D modal theory and controversial claims about the nature of phenomenal concepts. This lemma is interesting, but it's also complex and irrelevant to our discussion. Thus, for the purposes of this paper, I will ignore the issues that arise with respect to (5).<sup>43</sup>

## 2) Advancing the Dialectic

The 2-D argument is important because it uses new ideas about modality to advance the dialectic about the knowledge argument. This advance is easily illustrated: e.g., two of the most influential responses to earlier versions of the knowledge argument struggle to even challenge the 2-D argument.

### 2.1 Old Facts

One such objection is the popular old-fact/new-guise reply (e.g., Horgan 1984; Harman 1990; Loar 1997; Tye 2000).<sup>44</sup> This objection is built on the idea that we can know one and the same fact in different ways, i.e., under different guises. According to this objection, different propositions or modes of presentation can represent one and the same fact. Crucially, it's then argued that while Mary gains new phenomenal knowledge when she escapes her room—knowledge she could not acquire in her room—she does not thereby learn phenomenal facts of

---

<sup>43</sup> For more information, see Alter (2006, 2012), Chalmers (2002, 2004, 2009), and Pereboom (2011).

<sup>44</sup> This objection is one manifestation of the well-known Phenomenal Concept Strategy (also see: Sturgeon 1994; Papineau 2002; Block 2007; Ball 2009).

which she was previously unaware. Instead she gains new knowledge of a fact that she already knew in her room under a different guise. In other words, the claim is that she merely comes to know an old fact in a new way. The proffered explanation is that, upon experiencing colour for the first time, Mary acquires (or is enabled to deploy) new phenomenal concepts, and is able to gain new knowledge by use of these concepts. However, all of her new phenomenal concepts refer to the same properties and experiences as some of her pre-existing physical concepts. So she already knows all the phenomenal facts under a physical mode of presentation. Her escape into the coloured world merely prompts her to represent and know these facts in a novel, phenomenal way.

It's easy to see how the old-fact/new-guise reply challenges the traditional knowledge argument. The reply blocks the argument's inference from step 3 to 4:

3) Upon escape, Mary experiences colour for the first time and gains new knowledge about what it's like to see colour—i.e., she gains knowledge about the phenomenology of colour experiences.

4) Thus, upon escape, Mary learns new facts about the phenomenology of colour experiences.

In contrast, it's not clear that the old-fact/new-guise reply challenges the 2-D knowledge argument. This is because the old-fact/new-guise reply is designed to challenge the claim that Mary is ignorant of certain phenomenal facts while trapped. But the 2-D argument doesn't assert that Mary is ignorant of phenomenal facts. What this argument claims is that Mary cannot infer certain phenomenal truths (i.e., true propositions). So it seems that all the 2-D argument claims is that Mary cannot infer certain pieces of phenomenal knowledge. But this claim looks to be compatible with the old-fact/new-guise reply, for this reply states that there is some phenomenal

knowledge that Mary cannot acquire—and so, *mutatis mutandis*, cannot infer—until she escapes her room.

## 2.2 New Abilities

Another influential response that struggles to challenge the 2-D argument is the Ability Reply (Lewis 1990; Nemirow 2006; Cath 2009). This objection starts with a distinction between knowledge-that and knowledge-how. Knowledge-that is the typical province of epistemology: it's a mental state that takes a true proposition as its content. Knowledge-that is thus knowledge of truths. According to the ability reply, knowledge-how is different. The idea is that having knowledge-how is just matter of having an ability<sup>45</sup>. For instance, one piece of knowledge-how is the ability to dribble a basketball: one knows how to dribble just in case one has the ability to dribble. Crucially, it's supposed to follow from this account that know-how is *not* knowledge of truths or facts. This means that when one acquires a new piece of know-how, one does not thereby learn a new truth or fact. This claimed feature of know-how is the central ploy of the ability reply. For it leaves room for the ability reply to deny that Mary learns any new truths or facts when she escapes. The trick is that, by invoking knowledge-how, the ability reply can still accommodate the powerful intuition that Mary gains some sort of knowledge when she escapes. For the ability reply claims that Mary does indeed gain phenomenal knowledge, it's just that all she gains is knowledge-how—such as, e.g., the abilities to imagine, distinguish, and recognize colour experiences. This objection therefore argues that knowledge of what it's like to experience colour is a variety of know-how.

---

<sup>45</sup> Or, according to Cath (2009), having knowledge-how may be a matter of knowing a proposition under a “practical” mode of presentation. On this non-standard treatment of knowledge-how, the ability reply becomes a version of the old-fact/new-guise reply: the idea being that when Mary escapes, she gains new knowledge of old facts by deploying her newly acquired “practical” mode of presentation.

The ability reply straightforwardly challenges the traditional knowledge argument. The ability reply blocks the inference from step 3 to 4:

3) Upon escape, Mary experiences colour for the first time and gains new knowledge about what it's like to experience colour.

4) Thus, upon escape, Mary learns new facts about the phenomenology of colour experiences.

In contrast, it's not clear that the ability reply challenges the 2-D knowledge argument. The ability reply is designed to block the claim that Mary learns new phenomenal facts or truths when she escapes. But the 2-D argument doesn't focus on whether Mary gains knowledge when she escapes. It focuses on what Mary can and cannot infer while still in her room. And even if Mary will only gain know-how when she escapes, she still might be unable to infer certain phenomenal truths from the fundamental microphysical truth (P). If so, her inability will still entail that  $P \supset Q$  conditionals are not a priori. As such, it seems that the 2-D argument can still get going if—per the ability reply—Mary will not learn any phenomenal truths upon escape. I also can't see how the ability reply causes problems for any of the two-dimensional modal reasoning at work in premises (4) and (5). In short, the ability reply looks to be compatible with the key premises and inferences of the 2-D argument.

To be fair, it can still seem as if the ability reply conflicts with premise (1). This is because the ability reply appears to suggest that trapped-Mary can infer any relevant phenomenal truths. For if Mary only gains know-how upon escape, then it's natural to conclude that she already knows—and can thus infer from P—any relevant phenomenal truths. However, without additional argument, this inference is invalid. Mary can be ignorant of phenomenal truths in her room and yet merely gain know-how upon escape. Additional argumentation is needed to reach

the conclusion that, when she escapes, Mary not only gains nothing but new know-how, but also already knows all relevant phenomenal truths. The present ability reply does not include such an argument. Of course, it's open for someone to provide such argumentation, but my point about the struggles of the ability reply is already made. That is, while I won't rule out the possibility of an evolved Ability Reply 2.0, the present ability reply struggles to challenge the 2-D argument.

### 3) **~Infer-Q and Infer-M**

While the 2-D argument may avoid some extant objections, I still think that it fails to establish the falsity of physicalism. This is because, I argue, the 2-D argument relies on two theses that cannot both be true. These are:

**~Infer-Q-** In her room, Mary cannot infer phenomenal truths about experiences of colour from the fundamental microphysical truth.

**Infer-M-** In her room, Mary can infer every non-phenomenal macrophysical truth from the fundamental microphysical truth.

I first need to establish that the 2-D argument relies upon these twin theses. In the case of ~Infer-Q, this is an easy task. ~Infer-Q is just the first premise of the argument.

In the case of Infer-M, the 2-D argument relies upon it because the argument needs the 2-D thesis:

**2-D Thesis:** If T is not apriori, T is primarily contingent.

Remember, the 2-D Thesis is what allows the 2-D argument to move from premises about a priority and inference to conclusions about modality. The 2-D thesis is true only if Infer-M is true. That is, if Infer-M is false, then the 2-D thesis faces numerous counter-examples. The first reason is that, as everyone in the debate seems to think, the fundamental microphysical truth (P)

not only metaphysically but primarily necessitates ordinary, non-phenomenal macrophysical truths (Ms) (even my opponents accept this, e.g., Chalmers 2004, 2009).<sup>46</sup> This means that  $P \supset M$  conditionals are primarily necessary. With that in mind, note that a  $P \supset M$  conditional is a priori if and only if i) M is a priori, ii)  $\sim P$  is a priori, and/or iii) M can be inferred from P. But since non-phenomenal macrophysical truths are rarely a priori, and P is true *ex hypothesi*, it follows that, in effect,  $P \supset M$  conditionals are a priori if and only if Ms can be inferred from P. However, Ms can be inferred from P only if an idealized agent like Mary can so infer Ms. Therefore,  $P \supset M$  conditionals are a priori only if Mary can infer Ms from P—viz., only if Infer-M is true. It follows that if Infer-M is false, then  $P \supset M$  conditionals are both primarily necessary and not a priori. This means that if Infer-M is false,  $P \supset M$  conditionals are counterexamples to the 2-D thesis. As such, the 2-D thesis is true only if Infer-M is true.<sup>47</sup>

#### 4) An Overlooked Opportunity

As I just argued, the 2-D argument relies upon *two* foundational theses,  $\sim$ Infer-Q and Infer-M. The argument fails if either thesis is false. One might therefore expect both theses to come under fire from objections to the knowledge argument. But this is not the case. Infer-M has never been challenged. Some extant objections don't even address  $\sim$ Infer-Q and Infer-M, and those that do target only  $\sim$ Infer-Q. We already have two examples here. While the Ability and Old-fact/New-guise replies are influential, neither challenges Infer-M (as I said, it's hard to see

---

<sup>46</sup> While everyone in the debate seems to think that  $P \supset M$  conditionals are primarily necessary, no one actually argues for this conclusion. I try to remedy this situation in other work.

<sup>47</sup> It should be no surprise that some proponents (and sympathetic critics) of the 2-D argument also take great care to argue that ordinary, non-phenomenal macrophysical truths can be inferred from the complete microphysical truth (conjoined, perhaps, with T and I). For instance, see: (Jackson 1998, 2003b, 2005; Chalmers & Jackson, 2001; Chalmers 2009, 2012).



how these replies target any aspect of the 2-D argument).

I obviously want to change all this. For I'll soon argue that either Infer-M is false, or  $\sim$ Infer-Q is. But first, it helps to further survey extant objections to the knowledge argument. This will drive home the point that my objection takes a new tack in the dialectic.

#### **4.1 Incomplete Physical Knowledge Response (Aim: Falsify $\sim$ InferQ)**

This response argues that our marked microphysical ignorance leads us astray when we imagine Mary's situation (Stoljar 2006). The objection begins by pointing out that our scanty microphysical knowledge clearly does not equip us to infer phenomenal truths. It's then claimed that, in fact, our ignorance is such that we struggle to see how any amount of microphysical information might enable a person to infer phenomenal truths. It's therefore unsurprising that we presently intuit that Mary cannot infer relevant phenomenal truths from P. But this objection points out that it's a mistake to think that Mary's epistemic situation is anything like ours. By hypothesis, she has far more microphysical knowledge than we do; for she has *complete* microphysical knowledge. As such, the objection concludes, we must be open to the possibility that Mary can infer phenomenal truths from P.

While much remains to be said about the scope of microphysics and the possible extent of Mary's knowledge, one thing is clear: the incomplete-knowledge reply does not challenge Infer-M. It challenges  $\sim$ Infer-Q instead.

#### **4.2 The Indexical Reply (Aim: Explain away $\sim$ Infer-Q)**

The crucial idea here is that phenomenal knowledge is indexical knowledge, as phenomenal concepts are indexical concepts (Bigelow & Pargetter 1990; Ismael 1999; Perry 2001). The hope is that Mary's phenomenal ignorance is explained by this indexicality. This

objector basically argues that Mary, stuck in her room and armed only with P, is trapped in a context in which she cannot know certain phenomenal truths. Thus, Mary's phenomenal ignorance is supposed to be analogous to her inability to know many non-phenomenal indexical truths from her perspective. This result is supposed to show that Mary's phenomenal ignorance does not pose a problem for physicalism. Since her ignorance of other indexical truths is unproblematic, the same should go for phenomenal truths. In short, the indexical reply contends that Mary's phenomenal ignorance is but one instance of a widespread phenomenon that's compatible with physicalism.

Even this rough sketch of the indexical reply makes one thing clear: this reply does not target Infer-M. Instead, the goal is to argue that we can explain Mary's phenomenal ignorance (viz.,  $\sim$ Infer-Q) in a way that's friendly to physicalism.

#### **4.3 Acquaintance Reply (Aim: If anything, falsify $\sim$ Infer-Q)**

The Acquaintance Reply is a close cousin of the ability reply. Both begin with the idea that there is an important difference between ordinary knowledge-that, and some type of special knowledge that is not knowledge of truths or facts. The acquaintance reply draws a distinction between knowledge-that and acquaintance-knowledge (Conee 1994, Tye 2009). Acquaintance-knowledge is direct acquaintance with the thing that is the object of that knowledge. A subject thus has acquaintance-knowledge of a thing if and only if she is directly acquainted with that thing. The crucial claim is that acquaintance-knowledge is knowledge not of truths/facts, but of objects, individuals, or properties. This reply therefore argues that knowledge of what it's like to experience colour is a type of acquaintance knowledge. This means that when Mary leaves her room, the only thing she gains is direct acquaintance with certain phenomenal properties that

she's never encountered before.

The acquaintance reply does not target Infer-M. If it challenges the 2-D argument at all (my remarks about the ability reply apply here as well), it targets  $\sim$ Infer-Q. This is because if Mary gains only acquaintance-knowledge when she escapes, then perhaps a case can be made that she already knows—and can thus infer from P—any relevant phenomenal truths (see 2.2).

#### **4.4 Missing-Concept Reply (Aim: Block step 2 in the 2-D argument)**

This response starts with a key claim about the nature of phenomenal concepts. (Tye 2000; Hellie 2004; Montero 2007) This is that one can possess a given phenomenal concept  $Q$  only if one has had an experience with the phenomenal property picked out by  $Q$ . In short, the founding idea is that relevant conscious experiences are required to possess phenomenal concepts.<sup>48</sup> The missing-concept reply then argues that since Mary has never experienced colour, she cannot possess the phenomenal concepts that pick out the phenomenal properties of colour experiences. This reply then points out that since Mary lacks such concepts, it's no surprise that she's unable to infer the phenomenal truths in question. In short, the objection contends that Mary's inability to infer phenomenal truths is explained by her inability to possess phenomenal concepts (which is in turn explained by her paucity of colour experiences). The overall hope is that this explanation is compatible with physicalism.

Many details still need to be filled in here, but one thing is already clear. The missing-concept reply does not target Infer-M. Instead this reply attempts to explain away  $\sim$ Infer-Q.<sup>49</sup>

---

<sup>48</sup> As such, the Missing-concept reply belongs in the same family of objections as the old-fact/new-guise reply. Both are instances of what's called the Phenomenal Concept Strategy. See fn. 5.

<sup>49</sup> This strategy specifically manifests as an interesting objection to the inference from step 1 to step 2 of the 2-D knowledge argument.

1) Mary cannot infer certain phenomenal truths,  $Q$ , from P while still in her room.

2) Thus,  $Q$  cannot be inferred from P.

## 5) A New Objection

Once again, I focus on:

**~Infer-Q:** In her room, Mary cannot infer certain phenomenal truths, Q, from P (the complete microphysical truth).

**Infer-M:** In her room, Mary can infer all non-phenomenal, macrophysical truths, Ms, from P.

My initial target is Infer-M. But it's not a well-defined target, for as stated, Infer-M is multiply ambiguous. I'll consider each nearby precisification in turn. I'll then argue that, on each reading, either Infer-M is false, or the corresponding reading of ~Infer-Q is false. Either way, one thesis is false, and the 2-D argument is in trouble.

### 5.1 P Alone

Here's one reading of Infer-M:

**Infer-M1:** In her room, Mary can infer all Ms from P *alone*.

Infer-M1 claims that P provides Mary with all the information she needs to infer Ms. This means that, if Mary were only informed by P—i.e., if P was all the information she possessed—then she would still be able to infer Ms.

Infer-M1 is implausible. This is because in order to infer Ms, Mary must possess the requisite *M-concepts*—i.e., the ordinary, non-phenomenal macrophysical concepts involved with non-phenomenal macrophysical truths.<sup>50</sup> She needs these concepts to even entertain, let alone

---

This inference relies on the supposition that Mary is an idealized reasoner. But if Mary's situation prevents her from even possessing the phenomenal concepts involved with Q, then she's hardly ideal. In fact, even though she's very smart, her situation imposes a marked cognitive shortcoming. But if that's so, then her failure to infer Q does not entail that Q cannot be inferred from P. It might still be the case that truly idealized agents who possess phenomenal concepts can infer Q from P.

<sup>50</sup> I do not claim that she needs to possess the M-concepts in an expert way. For example, I do not here argue that in order to infer Ms, Mary needs to be able to explicitly define M-concepts. I just argue that she's able to infer Ms only if she possesses M-concepts to the extent that she's able to exercise them in thought.

infer, Ms. But, remember, P is just the fundamental microphysical truth.<sup>51</sup> As such, ordinary, non-phenomenal macrophysical concepts like *salt*, *elephant*, *crystalline*, and *mountain* do not feature in—and are not the topic of—P.<sup>52</sup> For such concepts are not part of the fundamental microphysical description of the world. So if Mary were informed solely by P, then not only would she lack any information about the use, origin, and meaning of M-concepts, but she also wouldn't grasp (i.e., know, believe and/or assume) any truths that directly involve these concepts as constituents. This impoverished cognitive situation looks to be incompatible with the possession of M-concepts. That is, it's plausible that if Mary were informed solely by P, she would not be informed enough to possess the M-concepts.<sup>53</sup> It follows that if Mary were solely informed by P, she would not be able to entertain, let alone infer, Ms. As such, when Infer-M is read as Infer-M1, Infer-M is implausible.

## 5.2 Enabling Information

Infer-M could also mean something like:

**Infer-M2:** In her room, Mary can infer all Ms when: i) P is her source of premises—i.e., of justifying information, and ii) she's also grasps some extra information, E, that enables her to possess the requisite M-concepts. Importantly, E only plays this *enabling* role. In no way does E justify or inform her inference of Ms.

Infer-M2 improves on Infer-M1 by accommodating the fact that Mary needs to possess the requisite M-concepts in order to infer Ms. To do this, Infer-M2 posits that in addition to P, Mary is also informed by some extra information, E, that enables her to possess M-concepts. The crucial idea here is that E does nothing more than enable her concept possession. This strategy

---

<sup>51</sup> Plus, perhaps, some locating information *I* and a totality clause *T*. See fn. 2.

<sup>52</sup> And the same goes for *I* and *T*.

<sup>53</sup> To be sure, though, nothing would prevent Mary from introducing or otherwise acquiring M-concepts. But to do so, she would need to expand her information base—viz., add to P. Thus, if Mary starts out being informed solely by P, and then acquires M-concepts, she ceases to be solely informed by P. She would then be informed by P and some extra information, E. I'll consider this possible scenario in subsequent sections.

preserves the claim that Mary can infer Ms from P. For even if P itself does not enable Mary to possess M-concepts, so long as P is sufficient to *justify* her inference of Ms (and it's possible for Mary to possess M-concepts), it's still true that Mary can infer Ms from P.

The basic idea of Infer-M2 is that, so long as Mary possesses the M-concepts, it's plausible that she can infer Ms from premises supplied by P. But notice that this does not require Mary to possess M-concepts in any sort of expert fashion. The idea is just that Mary needs to possess M-concepts in some way or other.

I argue that Infer-M2 is false. It's not the case that mere possession of M-concepts allows Mary to infer Ms from P. For we can think of examples where Mary possesses certain M-concepts, and yet cannot infer relevant Ms from P.

Suppose for instance that Mary can exercise the M-concepts *elephant* and *tusk* in thought, and is thus able to entertain conclusions involving these concepts. Also suppose that through her TV, phone, and books, Mary participates in a community of people who employ and are often well informed about *elephant* and *tusk*. But suppose that despite this social contact, Mary is deeply uninformed. While she's been provided with information beyond P to help enable her to possess *elephant* and *tusk*, this information is very limited. In fact, the only relevant truths she grasps (i.e., knows, believes and/or assumes) are the bland generalities *Elephants are a type of animal*, and *Tusks are a type of object*. She lacks any other beliefs or assumptions about and/or involving these concepts.<sup>54</sup> Despite her ignorance, in this scenario it's plausible (regardless of whether we're internalists or externalists about conceptual content) that Mary possesses the concepts *elephant* and *tusk*. Yes, she's markedly uninformed about them—perhaps we should

---

<sup>54</sup> Assume that Mary lacks information involving second-order concepts that refer to *elephant* and *tusk*.

say that she barely possesses these concepts—but they're nonetheless part of her conceptual repertoire.

Now further imagine that, consistent with Infer-M2, P is Mary's only source of justifying information. This means that all the premises Mary can draw upon to inform and justify her inferences are fundamental microphysical truths. So in this situation Mary both possesses the M-concepts *elephant* and *tusk* and can draw upon P as source of justifying information.

If Infer-M2 is true, then our version of Mary is in a position to infer macrophysical truths like *Elephants sometimes have tusks* (here I assume that her grasp of P provides her with the concept *have* and *sometimes*). But is this plausible? Is our version of Mary in a position to infer such an M? Intuitively, no, she is not. Her problem seems to be that while she possesses *elephant* and *tusk*, her limited information about these concepts prevents her from recognizing and understanding the connections between: i) these two M-concepts, ii) these M-concepts and relevant microphysical concepts, and iii) Ms involving these concepts and relevant microphysical truths. Without a sense of these connections, it's plausible that Mary cannot infer such Ms from P. That is, given her present grasp of *elephant* and *tusk*, P does not provide her with the justification needed to infer such truths.

This type of example falsifies Infer-M2. If Mary cannot infer Ms involving *elephant* and *tusk* from P in the above scenario, then it's not the case that she can infer Ms from P so long she possesses the requisite M-concepts. In other words, mere M-concept possession does not equip Mary to infer Ms from P.

### **5.3 Changing Cognitive Roles**

At this stage, I suspect some will protest that I gerrymander my version of Mary's

situation by restricting the set of extra information (E) that: i) transcends P, and ii) helps enable her to possess the M-concepts *elephant* and *tusk*. For the only pieces of extra, enabling information she grasps are *Elephants are a type of animal*, and *Tusks are a type of object*. But clearly we can change Mary's case such that she grasps more such information. Thus, I suspect some will protest that we can avoid my argument, and rescue Infer-M2, by simply providing Mary with more enabling information regarding M-concepts.

This protest is mistaken. If I'm right about my version of Mary's case, then my argument against Infer-M2 applies regardless of how we change the content of Mary's extra, enabling information, E. That is, if I'm right that *my version* of Mary cannot infer Ms from P, then, no matter how we change E's content, my argument shows that she cannot infer Ms from P. This is because since E merely plays an enabling role—as is stipulated by Infer-M2—so long as it fulfils this role, its content is irrelevant. For so long as a given E enables Mary to possess M-concepts, its content does not affect Mary's evidential or conceptual situation. For instance, given that E merely enables, my relatively uninformed Mary is in relevantly the same situation as another Mary with a more informative E. In both cases P is her sole source of justifying premises, and she's furnished with M-concepts by grasping some extra, enabling information, E. In both cases, she has the same intellect, the same premises, and the same conceptual competence. Given the lack of any relevant difference, our verdict about the cases should be the same. Thus, since my version of Mary cannot infer Ms from P, the same is true even when she is provided with more enabling information. This means that Infer-M2 is false no matter what quantity and type of enabling information we give to Mary. Mere possession of M-concepts does not equip her to infer Ms from P.



I should make one thing clear: with the above argument, *I do not reject* the idea that Mary might be able to infer Ms from P if, in addition to P, we provide her with a suitable set of extra information, E, regarding M-concepts like *elephant* and *tusk*. All I argue (so far) is that if E only serves to enable Mary to possess M-concepts, then it does not help Mary to infer Ms from P. So I'm still open to the possibility that Mary can infer Ms from P if, in addition to P, we provide her with an informative E, and allow E to do more than just enable Mary to possess M-concepts. In short, it remains to be seen whether we change Mary's inferential fortunes by changing E's cognitive role.

#### 5.4 Mastering M-Concepts

One suggestion on how to change E's cognitive role relies on the idea that there's an important difference between merely possessing a concept and *mastering* that concept. The basic thought is that mastering a concept involves being able to do things with it that go beyond the capabilities brought by mere possession. This suggestion proposes we provide Mary with extra information, E, that not only enables her to possess M-concepts, but also equips her to master M-concepts. This gives us a new version of Infer-M:

**Infer-M3:** In her room, Mary can infer all Ms when: i) P is her source of premises—i.e., of justifying information, and ii) she's also provided with some extra information, E, that equips her to not only possess but also *master* the M-concepts. Importantly, E only plays this 'mastering' role. In no way does E justify or inform her inference of Ms.

The change in E's cognitive role likely allows Infer-M3 to escape my previous argument, as a grasp of *elephants are a type of animal* and *tusks are a type of object*—and participation in a community of *elephant* and *tusk* users—is plausibly insufficient for mastery of these M-concepts. The change in E's cognitive role is also consistent with the crucial idea that E does not

play a justifying role for Mary. As before, this idea makes room for the claim that P is itself sufficient to justify Mary's inference of Ms.

At present, though, Infer-M3 is incomplete. Something needs to be said about what it is for Mary to master M-concepts. Only then can we understand what Mary's grasp of E is supposed to do for her, and therefrom determine whether Infer-M3 is true.

### 5.5 Mastery as Recognitional Skill

The literature contains three suggestions about what's involved with mastering a concept. I'll phrase these as sufficient conditions for mastery (and take their disjunction to be the necessary condition). The first suggestion says that:

**Recognitional Mastery:** Mary recognitionally masters a M-concept if she has the ability to accurately recognize things that fall into the extension of this concept, and thereby distinguish them from things that don't (e.g., Evans 1982; Peacocke 1992).

On this picture, Mary masters an M-concept like *grizzly bear* if she's able to recognize grizzly bears as such, and thereby sort grizzly bears from non grizzly bears.

If we understand concept mastery in this way, then Infer-M3 says that:

**Infer-M3\*:** In her room, Mary can infer all Ms when: i) P is her source of premises—i.e., of justifying information, and ii) she's also provided with some extra information, E, that equips her to not only possess M-concepts, but also *to accurately recognize things that fall into the extension of each M-concept*. Importantly, E only plays this 'mastering' role. In no way does E justify or inform her inference of Ms.

The problem is that Infer-M3\* looks false. There is no reason to think that Mary's new recognitional mastery enables her to infer Ms from premises solely supplied by P. Her relevant inferential capacities look to be the same as they would be if she merely possessed (i.e., did not master) M-concepts. Yes, it's true that she may now have a new and improved ability to learn all sorts of Ms by direct observation (e.g., for most any given bear, she can recognize and thus learn

whether it's a grizzly bear), but this does not suggest that she can infer Ms from P. In sum, if we understand concept mastery in a recognitional way, then Infer-M3\* is no improvement on Infer-M2: both theses look false.

## 5.6 Mastery as Inferential Skill

Another well-known suggestion is that:

**Inferential Mastery:** Mary inferentially masters a M-concept if she's disposed to draw and/or accept at least certain inferences that are justified by this concept's content (Brandom 2000, 2008; Peacocke 1992, 2008).

For instance, the content of the constituent M-concepts *poodle* and *dog* is what makes it the case that *Doug owns a dog* can be correctly inferred from *Doug owns a poodle*. The same conclusion cannot, for instance, be inferred from *Doug owns a tabby*. So the present suggestion might say that Mary masters *poodle* and *dog* if she's disposed to accept the above inference.

When applied to the task at hand, the notion of inferential mastery generates the idea that Mary masters M-concepts when she's disposed to infer at least some Ms from certain premise sets.<sup>55</sup> This makes Infer-M3:

**Infer-M3\*\*:** In her room, Mary can infer all Ms when: i) P is her source of premises—i.e., of justifying information, and ii) she's also provided with some extra information, E, that equips her to not only possess M-concepts, but also ensures that she masters M-concepts such that *she's disposed to infer* Ms from certain premise sets. Importantly, E only plays this 'mastering' role. In no way does E justify or inform her inference of Ms.

When Infer-M3 is construed in this way, it looks more plausible. To be sure, the fact that Mary is disposed to infer Ms from certain premise sets does not immediately entail that Mary can infer

---

<sup>55</sup> If one is disposed to draw inferences that are justified by the content of an M-concept, then one is both/either disposed to infer truths involving this concept—i.e., Ms—from certain premises, &/or to draw inferences from Ms. Only the former type of disposition is relevant here, as we're interested in whether inferential mastery of M-concepts enables Mary to infer Ms from P. Thus, in our debate, we should suppose—for the sake of discussion—that one inferentially masters an M-concept when one is disposed to infer Ms from at least certain premise-sets.

Ms *from* P, but one can see how a case might be made for Infer-M3\*\*. Since Mary is disposed to infer Ms from at least some premise sets, it's at least an open possibility she's disposed to infer them from P. It follows that we finally have a somewhat plausible reading of Infer-M.

However, let's not forget about ~Infer-Q. It's also a crucial part of the 2-D argument (in fact, ~Infer-Q is just the first premise of the 2-D argument).

**~Infer-Q:** In her room, Mary cannot infer certain phenomenal truths, Q (viz., phenomenal truths about what it's like to experience colour) from P.

This thesis is just as ambiguous as Infer-M. So if we understand Infer-M as Infer-M3\*\*, then we need to understand ~Infer-Q in the same fashion. For we cheat if we allow Mary to have all she needs to inferentially master M-concepts, but deny her corresponding mastery of phenomenal concepts. To be more precise, if we set up Mary such that she inferentially masters M-concepts, but does not inferentially master phenomenal concepts, then there's a problem with the 2-D argument's inference between:

- 1) Mary cannot infer certain phenomenal truths, Q, from P (the fundamental microphysical truth) while still in her room.
- 2) Thus, Q cannot be inferred from P.

If Mary does not master phenomenal concepts like she masters M-concepts, then she fails to fully master phenomenal concepts. But if she fails to fully master these concepts, then, plausibly, she's not a fully idealized agent. And if she's not a fully idealized agent, then her failure to infer Q does not entail that Q cannot be inferred from P. For all we know, it might be the case that fully idealized agents who inferentially master both M-concepts *and* Q-concepts can infer Qs from P.

The only way to block the above argument—and resist the call to interpret ~InferQ in the

same way as InferM— is to argue that it's impossible (at least for humans) to inferentially master phenomenal concepts. For instance, one might argue that we (even idealized subjects like Mary) can only hope to attain recognitional mastery of phenomenal concepts. This claim makes room for Mary to be fully ideal even if she inferentially masters M-concepts, but fails to likewise master phenomenal concepts. However, this line of thought is unmotivated and implausible. There is no reason to think that we cannot inferentially master phenomenal concepts. In fact, I suspect there's good reason to think the opposite: it seems like we're often disposed to draw or accept certain inferences that are justified by the content of phenomenal concepts. For instance, many philosophers are disposed to infer *My experience of the tomato is not phenomenally green* from *My experience of the tomato is phenomenally vermillion*, and many people are disposed to accept inferences like *The coffee does not taste delicious* from *The coffee tastes horrid*. Thus, there seems to be good evidence that we can and sometimes do inferentially master phenomenal concepts.

The result is that if we read Infer-M as Infer-M3\*\*, then ~Infer-Q becomes:

**~Infer-Q3\*\*:** In her room, Mary *cannot* infer relevant Qs (phenomenal truths about what it's like to experience colour) when: i) P is her source of premises—i.e., of justifying information, and ii) she's also provided with some extra information, E, that equips her to not only possess the requisite Q-concepts, but also ensures that she masters these concepts such that *she's disposed to infer* the Qs in question.

~Infer-Q\*\* is suspect. Or, to be more exact, ~Infer-Q\*\* looks to be just as implausible as Infer-M3\*\* is plausible. For in the same way that Mary's inferential mastery of M-concepts makes it more plausible that she can infer Ms from P, her inferential mastery of phenomenal concepts makes it more plausible that she can infer Qs from P. This is because since Mary is now disposed to infer Qs from at least some premise sets, it's at least an open possibility that she's

disposed to infer Qs from P. There is no reason why phenomenal truths and concepts should be different from Ms and M-concepts in this respect. The result is that Infer-M3\*\* is not a satisfactory reading of Infer-M. While this reading makes Infer-M plausible, it undermines the equally-important ~Infer-Q.

### 5.7 Mastery as Knowledge

The third and final suggestion reads as:

**Informed Mastery:** Mary masters a M-concept when she knows many truths about and/or involving this concept, and this knowledge not only enables her to possess this concept, but also (when applicable) justifies and informs her inferences.

The core idea here is that mastering an M-concept is a matter of putting knowledge about and/or involving this concept to work: by using it to deliberate, infer, and learn.

It might well be the case that Mary can infer Ms when she's supplied with P and has informed mastery of M-concepts. Nonetheless, this account of concept mastery cannot be used to formulate Infer-M3. For if we suppose that Mary needs to thus master M-concepts in order to infer Ms, then we forsake the idea that Mary can infer Ms *from P*. This is because if Mary thus masters M-concepts, then her supply of justifying premises includes not only P, but also an extra set of information, E, about/involving M-concepts. So on this account of concept mastery, Mary's knowledge of E plays not only an enabling but also a justifying role. As such, if Mary can infer Ms when she's supplied with P and relevantly masters M-concepts, all that follows is that she can infer Ms *from PE*. It doesn't follow that she can infer Ms from P. Quite the opposite. For if Mary needs to draw upon premises that transcend P in order to infer Ms, then she clearly cannot infer Ms from P.

In general, then, Infer-M3 is an unworkable version of Infer-M. Granting Mary mastery

of M-concepts does not ensure that she can infer Ms from P. Her mastery either contradicts Infer-M, does not enable her to infer Ms, or buys her the ability to infer Ms at the price of  $\sim$ Infer-Q.

## 6) Adding to Mary's Justifying Information

The prospects for Infer-M look bleak. Not only is Mary unable to infer Ms from P alone, but mere possession of M-concepts does not enable her to infer Ms from premises solely provided by P. We cannot even make headway by supposing that Mary not only possesses but also masters M-concepts. In light of these problems, the final option for a proponent of the 2-D argument is to run with the suggestion from the last section and forsake Infer-M. This person therefore admits that Mary cannot infer Ms from premises solely supplied by P. Instead he argues that in order to infer Ms, Mary needs to be provided with extra information, E, that not only enables her to possess M-concepts, but also serves as an additional supply of justifying information. Thus, the suggestion is that we provide Mary with an E that not only enables but also justifies her inferences.

But wait, doesn't the 2-D argument rely upon Infer-M? So doesn't this objector forsake the 2-D argument when he forsakes Infer-M? Not according to him. His idea is that we can rebuild the 2-D argument upon a new thesis:

**New Infer-M:** In her room, Mary can infer every ordinary, non-phenomenal macrophysical truth (every M) *from PE*.<sup>56</sup>

The aim is to replace Infer-M with the claim that Mary can infer all Ms from the conjunction of P

---

<sup>56</sup> More precisely: In her room, Mary can infer every M when: i) she's supplied with premises—i.e., justifying information—by not only P, but also by some extra information E, and ii) E also enables her to possess M-concepts.

and E.<sup>57</sup> To maintain consistency,  $\sim$ Infer-Q needs to be similarly updated. It becomes:

**New  $\sim$ Infer-Q:** In her room, Mary cannot infer certain phenomenal truths about what it's like to experience colour (certain Qs) *from PE*.<sup>58</sup>

The 2-D argument's first premise therefore becomes the claim that Mary cannot *even* infer phenomenal truths from PE. The structure of the argument is unchanged.<sup>59</sup>

- 1) In her room, Mary cannot infer certain phenomenal truths, Q, from PE.
- 2) Thus, Q cannot be inferred from PE.
- 3) Hence,  $PE \supset Q$  is not apriori.
- 4) If  $PE \supset Q$  is not apriori,  $PE \supset Q$  is primarily contingent. (From the 2-D thesis)
- 5) If  $PE \supset Q$  is primarily contingent,  $PE \supset Q$  is metaphysically contingent.
- 6) If physicalism is true,  $PE \supset Q$  is metaphysically necessary.
- 7) Therefore, physicalism is false.

## 6.1 A Futile Change

One cannot rescue the 2-D argument by switching from P to PE, and from the original  $\sim$ Infer-Q and Infer-M to their new formulations. This is true even if it turns out that New Infer-M and New  $\sim$ Infer-Q are both plausible. The problem is that the updated argument still relies upon the 2-D thesis.

**2-D Thesis:** If T is not apriori, T is primarily contingent.

And switching from P to PE does not save the 2-D thesis from counterexamples. This is because even if Mary can infer Ms from PE, this result does not challenge my argument that Mary is

---

<sup>57</sup> If you've been reading P as PTI, please note that E is supposed to be justifying information that also transcends I and T.

<sup>58</sup> More precisely: In her room, Mary cannot infer certain Qs when: i) she's supplied with premises—i.e., justifying information—by not only P, but also by some extra information E, and ii) E also enables her to possess Q-concepts.

<sup>59</sup> No additional changes needed. If P necessitates Q, then so does PE.



unable to infer Ms from P. In fact, the switch to PE is prompted by the seeming success of my argument. But if I've shown that Mary—*qua* idealized reasoner—cannot infer Ms from P, then I've also shown that  $P \supset M$  conditionals are not a priori. This is because, recall, a  $P \supset M$  conditional is a priori if and only if i) M is a priori, ii)  $\sim P$  is a priori, or iii) M can be inferred from P. And since non-phenomenal macrophysical truths are rarely a priori, and P is true *ex hypothesi*, it follows that, in effect,  $P \supset M$  conditionals are a priori if and only if Ms can be inferred from P.

$P \supset M$  conditionals are also primarily necessary (or so everyone seems to think). So my argument shows that  $P \supset M$  conditionals continue to be both primarily necessary and not apriori.  $P \supset M$  conditionals therefore continue to be counterexamples to both the 2-D thesis and the 2-D argument. As such, even if the switch to PE enables Mary to infer Ms, this result is futile, for it does not block my original objection to the knowledge argument.

## 6.2 Conclusion

The 2-D argument is the most sophisticated and compelling version of the knowledge argument. For one thing, by incorporating 2-D modal theory, the 2-D argument avoids some of the most popular responses to the knowledge argument. However, I argue that even the 2-D argument is unsuccessful. The problem is that it's built on two crucial theses,  $\sim \text{Infer-Q}$  and  $\text{Infer-M}$ , that are not both true. The root of the problem is that even if it's true that Mary cannot infer certain phenomenal truths, there's also good reason, I argue, to think that she cannot infer non-phenomenal, macrophysical truths. There is no inferential asymmetry between phenomenal and non-phenomenal truths. So Mary's seeming ignorance of phenomenal truths is not an exceptional phenomenon, and should not be seen as a significant point against physicalism.

## Chapter 4: Zombie Worlds and Strong Necessities

Philosophers of mind often find themselves in the strange and fortunate position of seriously debating the possibility of a zombie world. For as odd as it may sound to those outside the discipline, the issue of philosophical zombies directly bears upon the plausibility of physicalism as a theory of consciousness. If a zombie world is possible, physicalism is false. This is because—sadly—the zombies in question do not shamle, slobber or crave brains, and they are not covered with gashes and gore. Quite to the contrary, philosophical zombies are molecule-by-molecule physical duplicates of normally functioning humans. Thus, my zombie twin is a complete physical duplicate of me, with the same appearance, physiology and neurobiology. The only difference between zombies and normal humans is that zombies are not phenomenally conscious. Even though their neuro-sensory systems operate in the same way as ours, there is nothing it is like to be them, and they do not consciously experience the world as we do. A zombie world is likewise a physical duplicate of the actual world, the only difference being that phenomenal consciousness is totally absent in a zombie world. To paraphrase Robert Stalnaker (2002): the sun shines on the unfortunate and familiar-looking inhabitants of zombie worlds, but the lights are out in their minds.

Physicalism is incompatible with the possibility of a zombie world because this theory (or family of theories) has foundational modal commitments. In particular, physicalism is committed to the basic necessitation thesis:

**Fundamental:** Phenomenal truths are necessitated by some collection of physical truths.<sup>60</sup>

---

<sup>60</sup> Of course, Fundamental is just a placeholder for a more specific necessitation thesis, but it will serve for now.

This thesis is intended capture the physicalist idea that the world's physical structure and character somehow accounts for and gives rise to consciousness. This commitment is relevant because Fundamental is false if a zombie world is possible. For given that a zombie world is a physical but not phenomenal duplicate of the actual world, the possibility of such a world is thereby a direct counterexample to Fundamental. In summary then, because i) physicalism entails Fundamental and ii) the possibility of a zombie world falsifies Fundamental, it follows that physicalism is also false if a zombie world is possible.

Anti-physicalists therefore have much to gain by arguing that a zombie world is possible. This is typically done by offering a version of what is now referred to as the *conceivability argument*. Stripped to its essentials, the conceivability argument's core idea is that zombies are possible because they are conceivable. The structure of the argument is thus:

2. A zombie world is conceivable.
3. If a zombie world is conceivable, a zombie world is possible.
4. If a zombie world is possible, physicalism about phenomenal consciousness is false.

----

4. Therefore, physicalism about phenomenal consciousness is false.

This style of argument has a long and historically important dialectic. Philosophers have long objected to physicalist theories of the mind by arguing that since certain anti-physicalist scenarios are conceivable, they're possible. For instance, Descartes argued in this way that a thinking mind can exist without a body (Sixth Meditation, CSM II, 54). And Leibniz similarly argued—via his famous mill example—that cognition and perception are not physical phenomena (*Monadology*, 17, 1965). As one might expect, though, the conceivability argument

has since evolved in the face of various objections, and has become increasingly sophisticated. In particular, people like David Chalmers (1996, 2002, 2009) and Stephen Yablo (1993) have now devised versions of the conceivability argument that are designed to be immune to extant criticisms and counterexamples.<sup>61</sup>

In this paper I argue that even the most sophisticated versions of conceivability argument fail. In doing so I first chronicle some of the dialectical history of the argument and explain various modifications it has received in response to critiques. This makes it easier to explain present versions of the argument. For these arguments are complicated, and unless one knows some of the dialectical moves in play, it's hard to see the need for such complication. In any case, I then argue that while such contemporary arguments survive many challenges, they continue to be plagued by an old and serious problem: viz., it's highly doubtful that the possibility of a zombie world follows from its conceivability. I argue that despite the efforts made to bolster the conceivability argument against such criticism, there is still good reason to think that zombie worlds are impossible even if they are conceivable.

## **1. Clarifying the Discussion**

### **1.1 Zombies?**

More needs to be said about what we are supposed to be conceiving of when we attempt to conceive of a zombie world. A helpful way to gain clarity here is to first examine the details of the necessitation thesis physicalists are committed to. For as the possibility of a zombie world is supposed to falsify this thesis, a zombie world is helpfully understood as one that verifies this

---

<sup>61</sup> To be sure, some of these theorists have also devised new responses to such arguments—and thus, are not convinced by the arguments they proffer. Yablo is a notable example here.

thesis' antecedent but falsifies its consequent.

The standard consequent here is the conjunction of all phenomenal truths. Call it *Q*.

The standard antecedent here is the massive conjunction *PTI*. *P* is a complete collection of the fundamental microphysical truths (Chalmers & Jackson 2001; Chalmers 2002, 2009, 2012). *P* therefore specifies: i) the properties and relations of every (suitably) basic microphysical entity in the language of microphysics, and ii) the (suitably) fundamental laws governing the interrelations and behaviour of these microphysical entities. *T* is a 'that's all' clause that manifests as the truth *P is a complete statement of the microphysical truths and the world only features the entities, properties and laws that are needed to make P true* (Chalmers & Jackson 2001; Chalmers 2009, 2012). *T* is thus often abbreviated as the information *This world is a minimal P world*. Finally, *I* is a collection of contextual, locating information, often understood to be a conjunction of two truths of the form *I am A* and *Now is B* (Chalmers & Jackson 2001; Chalmers 2009, 2012). Here *A* is a description uniquely satisfied by some some subject and *B* is a description uniquely satisfied by some time at which the subject picked out by *A* exists.

The result is:

**Physicalist Necessitation:** *PTI* necessitates *Q*

Virtually all physicalists accept this thesis.<sup>62</sup> Indeed, it's often claimed that Physicalist Necessitation thesis is the minimal, foundational commitment of physicalism (e.g., Lewis 1994, Jackson 1998). Accordingly, recent proponents of the conceivability argument characterize a zombie world as one that falsifies *Q* (i.e., is somehow phenomenally distinct from the actual

---

<sup>62</sup> Strawson (2003, 2006) may be an exception, but it's not clear that he's a physicalist. I count him as a panpsychist—though, to be fair, he argues that physicalism entails panpsychism.

world) but nonetheless verifies PTI (i.e., is microphysically identical to the actual world) (Chalmers 2002, 2009). This gloss entails that when one tries to conceive of a zombie world, one tries to conceive of a world that verifies  $PTI \ \& \ \sim Q$ . It's then expedient to abbreviate by saying that when we attempt to conceive of a zombie world, we attempt to conceive of  $PTI \ \& \ \sim Q$ . The conceivability argument therefore becomes:

1.  $PTI \ \& \ \sim Q$  is conceivable.
2. If  $PTI \ \& \ \sim Q$  is conceivable,  $PTI \ \& \ \sim Q$  is possible.
3. If  $PTI \ \& \ \sim Q$  is possible, physicalism is false.
- 
4. Therefore, physicalism is false.

In what follows, I will assume this interpretation of the argument.

## 1.2 The Crucial Second Premise

The second premise of the conceivability argument is particularly important. It's this premise that allows the argument to move from a claim about the powers of human conception to, first, a modal conclusion and then on to an ontological conclusion about the nature of consciousness. Because it plays such a crucial role, this premise has attracted a great deal of debate. Importantly, much of this debate concerns the general thesis.

**CP:** If a scenario  $S$  is conceivable,  $S$  is possible.

This is because as the conceivability argument's second premise is but one instance of CP, the best way to motivate this premise is to defend CP. After all, if this principle is true, the conceivability of  $PTI \ \& \ \sim Q$  unquestionably entails this conjunction's possibility. For this reason, the debate about the conceivability argument's second premise is almost always cast as a debate

about CP.

In this paper I will argue that in even its most sophisticated form, CP is false. To do this I will first chronicle a series of important attempts to falsify CP, as well as a series of responses to these attempts. My chronicle will lead us to a current version of CP that is designed to be immune to all extant counterexamples. This refined thesis is my ultimate target.

## **2. A Complex Dialectic**

Further discussion largely turns upon how we should understand the sort of conceivability at hand. For even if we have some grasp of what we are trying to conceive when we attempt to conceive of a zombie world, much remains to be said about what's involved with this act of conception.

### **2.1 A priority**

The first thing that has often been noted about conceivability is that it is more closely tied to a priori rather than a posteriori reasoning. This is because it's clear that even if we can rule out a state of affairs by empirical investigation or resultant a posteriori reasoning, that state of affairs is often conceivable for us. For example, I can see that it is snowing outside and can thus rule out any state of affairs where snowflakes are not falling. Yet despite this fact, I can conceive of a situation where nature has blessed me with a sunny, clear day.

In stark contrast, it's intuitive that if I use a priori reasoning to rule out a state of affairs, then that state of affairs is not conceivable for me. Many examples support this thesis. To choose but one, I—and I know I am not alone here—cannot conceive of a world where  $2+2=5$ . Why? Well, plausibly because I know a priori that  $2+2=4$  and can thus a priori rule out any state

of affairs where  $2+2=5$ .

Examples like the ones listed above indicate that a scenario S is conceivable for a subject only if that subject cannot rule out S a priori. This entails that  $PTI \ \& \ \sim Q$  is conceivable in general only if it cannot be ruled out a priori—viz., only if  $PTI \supset Q$  is not a priori.

Given the intuitively tight connection between conceivability and a priori reasoning, it might also be argued that a zombie world is conceivable *if and only if* it cannot be ruled out a priori. This is a much stronger and more controversial claim, but it does leave us with a clarified conceivability argument, so let's start with it. The argument becomes:

1.  $PTI \ \& \ \sim Q$  is not ruled out a priori (i.e.,  $PTI \supset Q$  is not a priori).
2. If  $PTI \ \& \ \sim Q$  is not ruled out a priori,  $PTI \ \& \ \sim Q$  is possible.
3. If  $PTI \ \& \ \sim Q$  is possible, physicalism is false.

----

4. Therefore, physicalism is false.

## 2.2 Ideality

The above formulation is a good starting point for our discussion. This is because while the argument does much clarify our notion of conceivability, it also faces some obvious problems. First and foremost, the argument currently does not respect an important distinction between different sorts of conceivability. This is, as David Chalmers puts it, the distinction between *prima facie* and *ideal* conceivability (Chalmers, 2002). This distinction arises from the fact that, as the human intellect is limited and the demands of daily life heavy, oftentimes our a priori reasoning is casual and imperfect. With this point in mind, we might say that a scenario S is *prima facie conceivable* for a subject just in case S is not ruled out a priori by her casual,



everyday a priori reasoning. We can also generalize and say that  $W$  is *prima facie* conceivable in general if  $W$  is not ruled out by the a priori reasoning most subjects are routinely capable of. Thus,  $3 + 3 = 8$  is probably not *prima facie* conceivable, whereas a male barber who shaves all and only those who do not shave themselves perhaps is. Similarly, prior to Godel's work, an ordered list of the real numbers was certainly *prima facie* conceivable, and the same is probably true today.

A scenario  $S$  is *ideally conceivable*, on the other hand, just in case it cannot be ruled out by *ideal rational reflection* (Chalmers 2002). Hence, an ordered list of the real numbers is not ideally conceivable, as highly advanced a priori reasoning reveals that such a list is impossible. I am content here with the rough, intuitive idea that ideal rational reflection involves reasoning that cannot be defeated or improved upon. Ideal rational reflection is therefore not only mistake-free but also takes all relevant factors and evidence into consideration. Perhaps one can also say that ideal rational reflection is what is always exercised by an ideal reasoner; one that is free of contingent cognitive and/or environmental limitations.<sup>63</sup>

This distinction between *prima facie* and ideal conceivability is important because, at this stage, the first premise of the conceivability argument is ambiguous. We can read it as referencing either *prima facie* or ideal conceivability. This is problematic because *prima facie* conceivability clearly does not entail possibility. Many examples—including the ones I use above—attest to this fact. For instance, while a world containing a male barber who shaves all and only those who do not shave themselves may well be *prima facie* conceivable, such a world is not possible.

---

<sup>63</sup> See Chalmers 2002, 2009, 2012 for more on ideal reasoning and idealized agents.

Accordingly, we need to formulate the conceivability argument in terms of ideal conceivability. Then the argument need only rely upon the more plausible thesis:

**CP1:** If S is *ideally* conceivable, S is possible.

On this approach, PTI &  $\sim Q$  is conceivable iff PTI &  $\sim Q$  is not ruled out by ideal a priori reasoning—viz., if and only if PTI  $\supset Q$  is not ideally a priori. The resulting argument is:

1. PTI &  $\sim Q$  is ideally conceivable—viz., is not ruled out by ideal a priori reasoning.
2. If PTI &  $\sim Q$  is ideally conceivable, PTI &  $\sim Q$  is possible.
3. If PTI &  $\sim Q$  is possible, physicalism is false.

----

4. Therefore, physicalism is false.

### 2.3 Imagination

This reformulated argument still faces a noted class of counterexamples to CP1. These are examples featuring paired claims, both of which are not ruled out by ideal a priori reasoning, but only one of which is possible. Unproven and highly abstract mathematical hypotheses such as Goldbach's conjecture are typically to such examples. While a thesis like Goldbach's conjecture is either necessarily true or necessarily false, it's also plausible that neither Goldbach's conjecture nor its negation can be ruled out by ideal a priori reasoning. This is because the continual failure of mathematicians to prove or disprove this conjecture suggests that any determination of its truth is simply beyond (even idealized) human powers. The situation is therefore such that neither Goldbach's conjecture nor its negation are ideally a priori, yet only one is possible. This means that though the conjecture and its negation are both ideally conceivable, only one is possible. At least one of these hypotheses is therefore a counterexample

to CP1..

In order to combat this challenge, proponents of the conceivability argument introduce another distinction between two different types of conceivability. This is a distinction that has been characterized in multiple ways: Chalmers (2002, 2009) casts it as a divide between *positive* and *negative* conceivability, whereas Stephen Yablo (1993) describes it as a distinction between *conceivability* and *believability*. Regardless of the descriptions employed, though, negative conceivability/believability is just the sort of conceivability that we've discussed so far. That is, in unqualified form, a scenario S is negatively conceivable for a subject just in case she cannot rule it out a priori. So if we combine ideal and negative conceivability, we get the familiar result that S is ideally negatively conceivable iff S is not ruled out by ideal a priori reasoning.

The notion of positive conceivability adds on to the account of conception that we have developed so far. Here W is positively conceivable (or, in Yabloian terms, conceivable) for a subject if and only if she not only cannot rule out W a priori, but can also *robustly imagine* a scenario that verifies W (Chalmers 2002, 2009; Yablo 1993). So if we combine our distinctions once more, it follows that W is ideally positively conceivable iff one can robustly imagine a scenario verifying W in a way that survives ideal a priori reflection. The notion of robust imagination in play here is somewhat plastic. On one hand it can be understood to only include imagistic, quasi-perceptual imagining; where one imagines a scenario by creating a quasi-perceptual mental representation of it. This account entails that positively conceiving of a scenario involves forming a kind of picture (or soundscape, tastescape etc.,) of it in the mind's eye (or ear, or mouth, or....). Robust imagination can also be defined more broadly. For instance, some theorists speak of the “modal imagination” in this context (Chalmers 2002). In broad

strokes, one robustly modally imagines a scenario when one exercises one's imagination to get what seems to be a profound but non-imagistic sense of this scenario's nature.<sup>64</sup> The crucial idea here is that even without the mediation of images, it seems that we can sometimes gain what at least appears to be a deep grasp of an imagined scenario; a grasp that's richer than merely being unable to a priori rule it out (Chalmers, 2002). On this understanding, then, positive conceivability is distinguished from negative conceivability on the grounds that only the former involves employing one's imagination to acquire a robust sense of the seeming nature of the object scenario.

Examples are often used to help illuminate the elusive notion of robust modal imagination. For instance, it's plausible that we cannot quasi-perceptually imagine a world entirely composed of dark matter and dark energy. But given our knowledge of physics, it seems like we can still somehow profoundly grasp the nature of this scenario. In short, it seems that we can still imagine a dark-matter world in a way that richer than merely being unable to rule it out a priori. And if this the case, then on the view in question, I can not only negatively but also positively conceive of a dark-matter-world.

Regardless of how clear or convincing the distinction between positive and negative conceivability (or between conceivability and believability) may be, it's straightforwardly applied by some proponents of the conceivability argument. They use this distinction to bolster the argument against the apparent counterexamples posed by claims such as Goldbach's conjecture and its negation. The response is that while both of these hypotheses may well be

---

<sup>64</sup> I include "seems to be" here because this account of the modal imagination should not assume that a scenario must be possible in order to be robustly imaginable. This is not the main point of the account at hand. And as I argue later, there is reason to think that we can robustly imagine impossible scenarios.

negatively conceivable, neither of them is positively conceivable. In arguing this, proponents of the conceivability argument capitalize on the fact that their opponents already claim that the human intellect is unable to determine the truth value of such complex mathematical hypotheses. For if this is the case, then it's also reasonably plausible that we cannot exercise our imagination to get a robust grasp of a scenario that either falsifies or verifies this conjecture. When this point is combined with the seeming fact that we cannot quasi-perceptually imagine a scenario either verifying or falsifying Goldbach's conjecture, it follows that neither of these scenarios is robustly imaginable. And of course, if this is the case, then we cannot hope to positively conceive of the conjecture or its negation.

Proponents of the conceivability argument then proceed to formulate their argument in terms of positive conceivability. They thereby introduce the general thesis:

**CP2:** If S is ideally, *positively* conceivable, S is possible.

It is then claimed that PTI &  $\sim Q$  is ideally, positively conceivable. This means that PTI &  $\sim Q$  is:

i) not ruled out by ideal a priori reasoning (viz., that PTI  $\supset Q$  is not a priori), and ii) that a scenario verifying this hypothesis is robustly imaginable. The overall argument is then:

1. PTI &  $\sim Q$  is ideally, positively conceivable.
2. If PTI &  $\sim Q$  is ideally, positively conceivable, PTI &  $\sim Q$  is possible.
3. If PTI &  $\sim Q$  is possible, physicalism is false.

----

4. Therefore, physicalism is false.

## 2.4 Dimensions of Possibility

Despite its increased sophistication, this version of the conceivability argument is still

vulnerable to powerful counterexamples that cast doubt on its second premise. That is, there are a number of examples that seem to falsify CP2. These examples all involve the negations of necessary a posteriori truths such as *water is H2O*. This is because since a truth like *water is H2O* is a posteriori, its negation *water is not H2O* is both necessarily false and cannot be ruled out a priori, not even by ideal reasoning. What's more, it's also plausible that a world verifying *water is not H2O* is robustly and coherently imaginable. Take, for instance, the oft-discussed possible world that's exactly like the actual world except for the fact that XYZ and not H2O fills its oceans, lakes and streams. We certainly seem to robustly and coherently imagine this world when we wield it as a thought-example. So it seems that one only needs to imagine the XYZ world as the actual world to thereby imagine a world that verifies *water is not H2O*. For if the XYZ world is actual, water is XYZ.

It's therefore plausible that a world verifying *water is not H2O* is ideally, positively conceivable. But since *water is not H2O* is also necessarily false, it follows that this hypothesis is ideally, positively conceivable and yet impossible. *Water is not H2O* thereby serves as a direct counterexample to the claim that ideal, positive conceivability entails possibility.

To respond, proponents of the conceivability argument adopt a novel approach. Instead of further qualifying the sense in which a zombie world is conceivable, they introduce a distinction between two different sorts of possibility. This is the two-dimensionalist distinction between *metaphysical* and *primary* possibility.

## **2.5 Two-Dimensional Semantics**

Two dimensional theories of semantics first distinguish between primary and secondary intensions. Intensions are a kind of semantic value associated with representational items such

as sentences, thoughts, words, and concepts. In particular, intensions are functions from (centred) possible worlds to referents. The intension of an item therefore returns the referent (or lack thereof) of that item in any given (centred) possible world. The type of referent returned depends upon the type of representational item: intensions associated with sentences and thoughts return truth-values, whereas intensions associated with particular terms or concepts return particular entities or properties.<sup>65</sup>

The distinction between primary and secondary intensions arises because there are two different ways of considering a possible world when evaluating the referent of a representational item in that world. The first way one can do this is to consider the possible world as actual: that is, to suppose that this possible world is the actual world. But one can also consider a possible world as counterfactual: thereby acknowledging that the actual world is a certain way and considering the possible world in question only as an alternative way the actual world could have been. This distinction between different ways of viewing possible worlds is important because the referent we evaluate a representational item as having in a possible world can vary depending upon how that world is considered. For example, consider the referent of the term “water” in the XYZ world. Also assume (plausibly) that “water”: i) rigidly refers in all possible worlds, and ii) that it functions to pick out the actual substance that has certain watery functional, chemical and appearance properties: such as being liquid between 0 and 100 °C, having surface tension when liquid, being colourless and being able to quench flames. Now, when we consider the XYZ world as actual, “water” picks out XYZ in that world. This is because when we consider this world as actual, XYZ is being taken to have all the watery properties in the actual world. On the

---

<sup>65</sup> For an excellent overview of this material, see Chalmers 2006.

other hand, when we consider the XYZ world as counterfactual, “water” continues to pick out H<sub>2</sub>O there.

Accordingly, the truth value we assign of a thought in a given world can also vary depending upon whether we consider that world as counterfactual or actual. *Water is H<sub>2</sub>O* is a clear example here. When we consider the XYZ world as actual, *water is H<sub>2</sub>O* comes out false in that world. For if the XYZ world is the actual world, *water* picks out not H<sub>2</sub>O but XYZ (and, indeed, does so rigidly in all possible worlds). However, when we consider the XYZ world as counterfactual, *water is H<sub>2</sub>O* is true. For *water* rigidly refers in all possible worlds, and what it rigidly refers to is determined by its referent in the acknowledged actual world—which is H<sub>2</sub>O. When we consider the XYZ world as counterfactual, it does not matter what substance plays the watery-role in that world.

The system of distinct primary and secondary intensions is designed to capture the apparent fact that the referent we evaluate an item as having at a world depends upon how this world is considered. The primary intension of an item A returns the referent of A in any given possible world considered as actual. The secondary intension of A returns the referent of A in any given possible world considered as counterfactual. The primary intension of *water is H<sub>2</sub>O* therefore returns the truth value 'true' in the XYZ world, whereas the corresponding secondary intension returns 'false'.

## **2.6 Two-Dimensional Modality**

Our distinction between types of intensions allows us to also distinguish between different notions of possibility: viz., between different ways for a sentence or thought to be



possible. First we have the more familiar notion of metaphysical or secondary possibility.<sup>66</sup> Here a claim T is metaphysically possible just in case its secondary intension returns 'true' in some possible world. That is, T is metaphysically possible just in case it is true in some world considered as counterfactual. On the other hand, T is primarily possible just in case its primary intension returns 'true' in some possible world. This means that T is primarily possible just in case it is true in some possible world considered as actual. Notions of primary and secondary necessity and contingency can then be easily defined in a similar fashion. For instance, we can say that T is primarily necessary just in case T is true in all possible worlds considered as actual.

Proponents of the conceivability argument put these dimensions of possibility to work by arguing that while ideal, positive conceivability may not entail metaphysical possibility, such conceivability does entail primary possibility.

**CP3:** If S is ideally, positively conceivable, S is *primarily* possible (Chalmers 2002, 2009).

This position is founded on the idea that the negations of a posteriori necessities such as *water is not H<sub>2</sub>O* are primarily possible even though they are not metaphysically possible. This idea is highly plausible. There is excellent reason, for instance, to think that *water is not H<sub>2</sub>O* is true in some world considered as actual. Take the aforementioned XYZ-world as an example. If this world is the actual world, then not only is water actually XYZ, but *water* picks out XYZ. As a result, if the XYZ-world is actual, *water is XYZ* is true. The XYZ-world therefore is but one world that plausibly sustains the primary possibility of *water is not H<sub>2</sub>O*.

## 2.7 A Two-Dimensional Argument

---

<sup>66</sup> Chalmers (2002, 2009) speaks of secondary possibility, but it is clear that secondary and metaphysical possibility are the same thing. I simply prefer the more common terminology.

The conceivability argument now becomes (Chalmers 2002, 2009):

1. PTI &  $\sim$ Q is ideally, positively conceivable.
2. If PTI &  $\sim$ Q is ideally, positively conceivable, PTI &  $\sim$ Q is primarily possible.
3. If PTI &  $\sim$ Q is primarily possible, PTI &  $\sim$ Q is metaphysically possible. (synecdochic premise)
4. If PTI &  $\sim$ Q is metaphysically possible, physicalism is false.

----

- 5) Therefore, physicalism is false.

This version of the conceivability argument needfully contains a new step: premise 3.

This step rests upon subtle arguments about the relationship between the primary and secondary intensions of phenomenal concepts.<sup>67</sup> While this additional argumentation is interesting, I will set it aside in this paper. It's not my topic. I continue to focus on premise 2.<sup>68</sup>

### 3. The State of the Opposition

#### 3.1 Strong Necessities

As I said before, the second premise of the conceivability argument is cast as an application of the general thesis:

**CP3:** If S is ideally, positively conceivable, S is primarily possible.

---

<sup>67</sup> See: Chalmers 2002, 2009; Pereboom 2011.

<sup>68</sup> Chalmers (2002, 2009) has an even more complicated version of the conceivability argument. It is:

1. PTI &  $\sim$ Q is ideally, positively, **primarily** conceivable.
2. If PTI &  $\sim$ Q is ideally, positively, **primarily** conceivable, PTI &  $\sim$ Q is primarily possible.
3. If PTI &  $\sim$ Q is primarily possible, PTI &  $\sim$ Q is metaphysically possible **or Russellian Monism is true**.
4. If PTI &  $\sim$ Q is metaphysically possible, physicalism is false.

----

5. Therefore, physicalism is false **or Russellian Monism is true**.

In this paper I will ignore these additional wrinkles because they do not substantially strengthen the conceivability argument, are irrelevant to my eventual arguments, and require lengthy investigation.

Objections to the second premise are therefore attempts to provide convincing counterexamples to CP3. Counterexamples need to be hypotheses that are ideally, positively conceivable and yet primarily *impossible*. This means that such a counter-example needs to be: i) not ruled out by ideal a priori reasoning, ii) verified by a robustly imaginable scenario, and yet iii) falsified by all possible worlds considered as actual. I call such falsehoods *impossible dreams*.

Strangely, though, the dialectic only indirectly discusses impossible dreams. For some reason most writers directly focus on the issue of whether so-called *strong necessities* obtain. Strong necessities—if they obtain—are the negations of impossible dreams. Hence, a strong necessity is any truth that's primarily necessary and yet ideally, positively conceivable as false. This means that a strong necessity is:

- i) True in all possible worlds considered as actual, ii) falsified by a robustly imaginable scenario, and iii) not a priori.

While the focus on strong necessities is somewhat odd, it still connects to the central issue. If strong necessities obtain, their negations do too—i.e., impossible dreams do too. And if impossible dreams obtain, then CP3 is false.

The debate over CP3 is consequently about strong necessities. This debate is vast and complex. Many putative strong necessities have been proposed. Notable examples include<sup>69</sup>:

- 1) A posteriori identities where the expression on each side of the identity picks out the referent by virtue of one of its essential properties (there are hints of this in Block 2006).
- 2) Identities involving perceptual demonstratives, such as “that experience = this experience” (Schiffer 2003).
- 3) Laws of nature (e.g., Shoemaker 1998).
- 4) Existence claims about necessary existents (Yablo 1999). For example, suppose that

---

<sup>69</sup> In making this survey, I'm indebted to Chalmers 2009. I cannot improve upon his thorough map of the literature.

Wally is a toaster who exists necessarily if he exists at all. The idea is that *Wally exists* and *Wally does not exist* are therefore both candidate strong necessities.

5) Material conditionals involving PTI and non-phenomenal macrophysical truths like *granite is an igneous rock*. That is, material conditionals of the form  $PTI \supset M$  (where M stands for any non-phenomenal macrophysical truth) (Block & Stalnaker 1999).

#### 4) **PTI $\supset$ M Conditionals**

While all of these proposed counterexamples raise fascinating issues, I cannot discuss them all here.<sup>70</sup> I will instead focus on what I take to be the most promising suggestion: number (5). In doing so, I will provide a new argument for the conclusion that at least some  $PTI \supset M$  conditionals are strong necessities. So on my view, examples of strong necessities might include: *PTI  $\supset$  granite is an igneous rock* and *PTI  $\supset$  elephants have trunks*. I thereby argue that at least some  $PTI \supset M$  conditionals are: i) primarily necessary, ii) not a priori, and iii) falsified by robustly imaginable scenarios.

##### 4.1 **Primarily Necessary**

I won't spend time arguing that  $PTI \supset M$  conditionals are primarily necessary. No arguments are needed. Everyone in the debate already accepts that  $PTI \supset M$  conditionals are primarily necessary. Even Chalmers (e.g., 2009, pp. 22) admits that this is true.

##### 4.2 **Not A Priori**

As with any material conditional, there are three mutually inclusive ways for a  $PTI \supset M$  conditional to be a priori. First,  $PTI \supset M$  is a priori if M is a priori. Second,  $PTI \supset M$  is a priori if  $\sim(PTI)$  is a priori. Finally,  $PTI \supset M$  is a priori if M is a priori inferable from PTI. However, despite this range of options, it is clear that if  $PTI \supset M$  conditionals are universally a priori, then

---

<sup>70</sup> See Chalmers (2009) for discussion of the other proposals.

this is due to the fact that every  $M$  (i.e., every non-phenomenal macroscopic truth of the actual world) is a priori inferable from PTI. This is because the first two options are simply not available in most instances. For one thing, fundamental microphysical *truths* are clearly not a priori falsifiable. For another, the vast majority of non-phenomenal macroscopic truths are not independently a priori. We can thus determine whether  $PTI \supset M$  conditionals are a priori by determining whether we (or a least idealized versions of us) can a priori infer  $M$ s from PTI.

In a previous chapter I explore the issue of whether we can a priori infer  $M$ s from PTI—see “Metaphysical Explanation and Inference”. My conclusion is that we cannot. Not even ideal agents can infer  $M$ s from premises supplied only by PTI. On the strength of these arguments, I conclude that  $PTI \supset M$  conditionals are not a priori.

### 4.3 Imaginable as False

A scenario falsifying a  $PTI \supset M$  conditional (that is, a  $PTI \& \sim M$  scenario) is robustly imaginable just in case we can “modally” imagine this scenario in a way that's richer than merely being unable rule it out a priori. When we imagine a scenario in this way, we gain a seemingly profound grasp of the object scenario's nature. So a  $PTI \& \sim M$  scenario is robustly imaginable just in case we can gain an apparently profound grasp of a scenario that's microphysically identical to and yet macrophysically different from the actual world (and that also verifies  $I$  and  $T$ ). So, for example, if we can profoundly grasp the nature of a scenario that verifies PTI and yet contains a living being in Lenin's tomb, it follows that a scenario falsifying  $PTI \supset \text{Lenin is dead}$  is robustly imaginable.

With the issue is framed in this way, we have good reason to think that at least some  $PTI \supset M$  conditionals are falsified by robustly imaginable scenarios. This is because it at least seems

to be the case that we can robustly imagine many PTI &  $\sim M$  scenarios. For instance, it seems to me that I can robustly imagine a scenario that's microphysically identical to the actual world, but which is also such that all DeLoreans are capable of time-travel. It also seems like I can hold PTI fixed and robustly imagine that DeLoreans can be used to *change* the past.<sup>71</sup> As such, this imaginable scenario jointly verifies PTI and *DeLoreans can be used to change the past* (D). But of course, DeLoreans cannot be used to change the past: this is an impossible feat. This entails that the material conditional composed of PTI and the non-phenomenal macrophysical truth *DeLoreans cannot be used to change the past* ( $\sim D$ ) is true—in fact, necessarily so. But now we have a striking example, for while  $PTI \supset \sim D$  is true, we can robustly imagine the falsifying scenario PTI & D.  $PTI \supset \sim D$  is thus a prime example of a  $PTI \supset M$  conditional that's robustly imaginable as false.

There are other examples too. For instance, it seems to me that I can robustly imagine a scenario which is microphysically identical to the actual world, but which is also such that individuals in distinct possible worlds can detect each other speaking. In this scenario, occupants of other possible worlds can perhaps sense me singing in the shower. All I have to imagine is, first, a duplicate of the actual world, and, second, another world where people miraculously detect my voice whenever I warble “Ol’ Man River”. This scenario jointly verifies PTI and *Other-world individuals detect my singing* (S). The problem here is that other-world individuals clearly do not detect my singing. In fact, it's probably impossible for this to occur. Even modal realists hold that possible worlds are causally and spatio-temporally isolated from each other (e.g., Lewis 1986). It follows that the material conditional composed of PTI and the non-

---

<sup>71</sup> Peter Kung (2010) is a fellow traveller here. We share the same (excellent) taste in movies and thought-examples.

phenomenal macroscopic truth *Other-world individuals do not detect my singing* ( $\sim S$ ) is true—in fact, necessarily so. So once again we have a telling example:  $PTI \supset \sim S$  is true even though we can robustly imagine the falsifying scenario  $PTI \ \& \ S$ . It follows that we have good evidence to conclude that we can robustly imagine at least some  $PTI \ \& \ \sim M$  scenarios.

Possible Responses:

**1) *Wait, these scenarios are impossible!*** This objection contends that if a scenario is impossible, then it fails to be robustly imaginable. It then points out that scenarios jointly verifying  $PTI$  and, respectively, *Other-world individuals detect my singing* ( $S$ ) and *DeLoreans can be used to change the past* ( $D$ ) are impossible. This is because  $D$  and  $S$  are themselves metaphysically impossible.

In response, while I understand that it's a bit odd to think of robustly imagining an impossible scenario, it's clear that we can perform this feat. We often use our faculties to imagine scenarios that are not only impossible, but which we know to be impossible. Even close variations of my previous examples can be used to illustrate this fact. For instance, as my fellow traveller Peter Kung argues:

Whatever our familiar creative faculty is, it certainly seems that we can use it to imagine a teenager travelling back to 1955 in a DeLorean and, through a series of mistaken-identity-fueled madcap adventures, *changing* his father from ineffectual loser into confident leader. (Kung 2010, pp. 2)

Kung is clearly right that we can imagine such a story. I think it's clear that if a movie depicts a given scenario, then that scenario is robustly imaginable. But as Kung also points out, our young hero's feat is well known to be impossible. One cannot change the past. So despite the fact that our hero's time-travelling adventure is robustly imaginable, it is impossible.

Another example can be drawn from contemporary metaphysics. Consider the various

answers to the special composition question: we have Mereological Nihilism, Universalism and several theories of restricted composition. Nihilism holds that composition never occurs (viz., it's never the case that some Xs compose a Y), Universalism holds that composition always occurs (viz., any collection of Xs composes a Y), and restricted views put forth some intermediate position. While many of these theories are implausible, it's intuitive that all of them are robustly imaginable. In other words, we can robustly imagine a nihilistic world, a universalistic world and so forth. To be sure, these worlds are bizarre, but they do not surpass our powers of imagination.

While each of these scenarios is imaginable, it seems that only one is possible. This is because there's good reason to think that conditions of composition do not vary between possible worlds. In short, it's plausible that if an answer to the special composition question is actually true, then it's necessarily true.<sup>72</sup> Since the various answers are incompatible, it follows that all but one of them are necessarily false. Consequently, all but one of these theories are impossible and yet robustly imaginable.

In conclusion, these examples indicate that this objection relies on a false principle. It's simply not the case that impossibility entails unimaginability. Hence, it's not a problem that scenarios verifying *Other-world individuals detect my singing* and *DeLoreans can be used to change the past* are impossible.

**2) *These Scenarios are not Ideally Imaginable:*** This objection begins by pointing out that in asking whether  $PTI \supset M$  conditionals are robustly imaginable as false, we're surely asking whether they're ideally imaginable as such. After all, this question is part of the larger issue of

---

<sup>72</sup> Cameron (2007) is the only proponent of contingent composition that I know about.



whether ideal, positive conceivability entails primary possibility. We thereby need to think about whether an idealized version of our creative faculty permits us to robustly imagine PTI &  $\sim$ M scenarios. This objection contends that even if such scenarios are *prima facie* imaginable, they do not survive the ideal imagination.

In response, while I can understand being cautious about the capacities of our ideal imagination, I see little particular reason to doubt that such scenarios are ideally imaginable. On this score, I can only think of the argument that since we're not idealized 'imagers', and since PTI &  $\sim$ M scenarios seem pretty weird, we should be wary of proclaiming them ideally imaginable. Perhaps this is true, but a reason to be wary is not the same thing as a reason for active doubt. So until we're supplied with specific reason to think that our scenarios fail to be ideally imaginable, we can (warily) follow our intuitions to the opposite conclusion.

Moreover, my opponents cannot effectively use this sort of objection in the larger debate over the conceivability argument. This is because if such reasoning casts doubt on the robust imaginability of PTI &  $\sim$ M scenarios, it casts the same doubt on the robust imaginability of PTI &  $\sim$ Q scenarios. That is, I can equally use this objection to argue that zombie worlds fail to be robustly imaginable. For since zombie worlds are also very weird scenarios, we should likewise be wary about proclaiming them to be ideally, robustly imaginable. But if zombie worlds fail to be robustly imaginable, then they also fail to be positively conceivable; and the conceivability argument never gets off the ground. Proponents of the conceivability argument therefore cannot propose this objection without also endangering their own argument. In their hands, this objection is a *tu quoque*.

#### **4.4 Conclusion**

We now have good reason to judge that at least some  $PTI \supset M$  conditionals are strong necessities: viz., that some such conditionals are i) primarily necessary, ii) not a priori, and iii) robustly imaginable as false. The negations of these conditionals—i.e.,  $PTI \ \& \ \sim M$  conjunctions—are thus ideally, positively conceivable and yet primarily impossible. This means that  $PTI \ \& \ \sim M$  conjunctions are counterexamples to the claim that ideal, positive conceivability entails primary necessity. Such conjunctions therefore show us that even when update and two-dimensionalize our notions of conceivability and possibility, it's still not the case that the conceivability of a zombie world entails its possibility.

## **5. A Response from Modal Theory**

The most straightforward objections to my argument take issue with some horn of my argument for the strong necessity of  $PTI \supset M$  conditionals. In particular, objectors like Chalmers and Jackson may want to criticize my arguments for the non-a priori of such conditionals (see, e.g., Chalmers & Jackson 2001, Chalmers 2012). I will not discuss such responses here. I already cover such ground in other papers—see “Mary’s Macrophysical Ignorance” and “Grounding and Inference”. In this paper I want to instead focus on a more general objection to the idea that strong necessities exist. For in addition to attacking proposed examples of strong necessities, Chalmers offers a positive argument for their non-existence; one that makes claims about the overall shape of our best modal theory. As such, I want to now consider whether this argument spawns any general considerations against strong necessities that might override my arguments.

### **5.1 Modal Monism**

Chalmers' overall idea is that we should reject strong necessities because they're incompatible with our best overall modal theory (1999, 2009). To develop this suggestion, he starts with the claim that defenders of strong necessities must hold that the space of possible worlds does not coincide with the space of ideally, positively conceivable (henceforth, IP conceivable) scenarios. For by defending strong necessities, one claims that there are certain IP conceivable scenarios that are neither primarily nor metaphysically possible. The defender of strong necessities therefore endorses a view that is sometimes referred to as **modal dualism**, as she holds that there are two distinct ranges of scenarios at which we can seemingly evaluate the truth value of a hypothesis (Chalmers 1999, 2009). First there is the range of IP conceivable scenarios, second there is the range of genuinely possible worlds, and these these ranges are not identical—as certain scenarios outstrip the range of possible worlds.

Chalmers then contrasts modal dualism with the opposing view: modal monism. Monism is the theory that the range of possible worlds perfectly coincides with the range of IP conceivable scenarios, leaving us with only one space of cases at which to evaluate the truth value of a hypothesis (Chalmers, 1999, 2009). On this view, a world is possible just in case it is ideally, positively conceivable. Modal monism is therefore patently incompatible with the existence of strong necessities.

Having introduced this distinction between monistic and dualistic modal viewpoints, Chalmers argues that monism is a better overall theory. He does this by first pointing out that monism is less complicated and introduces fewer classes of phenomena. For while dualism posits separate spaces of IP conceivable scenarios and possible worlds, monism collapses these two spaces into one plenum.

With parsimony in hand, Chalmers then argues that monism accounts for all modal phenomena that we need to (and want to) invoke possible worlds to explain. That is, he argues that we can use this single space of ideally, positively conceivable worlds/scenarios to explain—among other things—modality, rational inference, counterfactuals, semantic content and the content of thoughts. Chalmers then invokes considerations of theoretical simplicity. His idea is that we should prefer monism because it's equally effective and more economical. The argument is basically this:

- 1) Modal monism accounts for all modal phenomena and supports any explanation or theory that invokes possible worlds.
- 2) Modal monism is simpler than modal dualism and invokes fewer classes of entities.
- 3) Therefore, modal monism should be preferred over modal dualism. I.e., modal monism is our best modal theory.
- 4) By definition, modal monism rules out the existence of strong necessities.
- 5) Therefore, our best modal theory rules out strong necessities, and, so, we should deny their existence.

#### Response 1

I argue that Chalmers' objection begs the question. His first premise is to blame. For it's logically equivalent to the claim that strong necessities do not exist. Two points are important here. First, by definition monism cannot account for—in fact, cannot abide—the existence of strong necessities. Second, we must remember that if strong necessities exist, their existence is a modal phenomenon. So if strong necessities exist, monism cannot account for all modal

phenomena. And if monism accounts for all phenomena, then strong necessities do not exist. Chalmers' first premise therefore amounts to little more than the claim that strong necessities do not exist. Chalmers consequently begs the question with his subsequent argument. For he essentially infers the non-existence of strong necessities from the claim that they do not exist.

To fairly adjudicate whether modal monism is a satisfactory modal theory, we need to *first* figure out whether strong necessities exist. That is, we cannot use conclusions about the completeness of modal monism *qua* modal theory as a way to draw conclusions about the possibility of strong necessities. We have to reason in the other direction.

### Response 2

Despite the above response, one might still worry that we should set aside strong necessities when comparing modal dualism and monism. This is because one might worry that strong necessities are overly partisan and controversial as examples of modal phenomena that monism cannot explain. This is a natural thought (though I disagree, as I actually have an argument that strong necessities exist), so let's set strong necessities aside for the moment. I'm content to do this because I do not need strong necessities to argue that modal monism cannot account for all modal phenomena. There are other modal phenomena that are incompatible with monism. In particular, there are possible worlds which do not count as such on monism. The worlds in question—call them *alien worlds*—feature *alien properties*. Alien properties are properties that are neither instantiated by any actual entity nor composed of any properties instantiated by actual entities. Alien properties therefore do not even in part obtain in our world. Despite their absence in the actual world, it's intuitive that alien properties are possible (see, e.g., Lewis 1986). This is because it's intuitive that our world could have featured properties

completely different from those that actually obtain.<sup>73</sup> In other words, why think that every possible world is 'built out of' the elements of this one?

For our purposes, the pivotal feature of alien properties is that they fail to be robustly imaginable. After all, these properties are wholly unlike any actual properties. So not only are we unable to call up an 'image' of such properties before our mind's eye (or ear, or tongue), but we also cannot exercise our imagination to gain a profound grasp of the nature of such properties. Alien properties are completely beyond our ken. It follows that alien worlds also fail to be robustly imaginable. After all, if we cannot robustly imagine the properties that obtain in a world, we cannot completely imagine this world. The best we can do is gain a profound grasp of the partial, non-alien nature of this world, and that is not enough to robustly imagine the entire world. In any case, what's important is that since alien worlds are not robustly imaginable, they also are not positively conceivable. For remember, a world is positively conceivable only if we can robustly imagine it.

The result is that alien worlds are possible and yet fail to be ideally, positively (IP) conceivable. This result causes serious trouble for modal monism, for it states that the range of IP conceivable worlds is not the same as the range of possible worlds. That is, modal monism is incompatible with the possibility of alien worlds. It follows that modal monism cannot account for all plausible modal phenomena. Indeed, monism seemingly can only hope to account for

---

<sup>73</sup> That being said, I admit that a number of theories of possible worlds rule out the possibility of alien worlds—Armstrong's (e.g., 1989) combinatorial theory being perhaps the most notable example. The existence of such theories indicates that we can be led to reject the possibility of alien worlds. Still, the fact that such theories are incompatible with alien worlds is typically regarded as one of their shortcomings. For instance, Lewis argues that one of the main advantages of his modal realism is that it licences the possibility of alien worlds (Lewis, 1986). The presence of theories such as Armstrong's is compatible with us having strong *prima facie* reason to think that alien worlds are possible.

modal phenomena associated with **non-alien** worlds.

It's worth pointing out that alien worlds trouble modal monism in a novel way. Alien worlds argue for distinct ranges of IP conceivable scenarios and possible worlds by indicating that there are possible worlds that do not correspond to conceivable scenarios. Alien worlds therefore suggest that, in one dimension, the range of possible worlds outstrips the range of IP conceivable scenarios. This is just the opposite of strong necessities. These truths show that there are conceivable scenarios which fail to be possible. So if we both admit the possibility of alien worlds and accept the existence of strong necessities, we are forced to modal dualism from two directions. On one hand there are possible worlds that are not conceivable, and on the other there are conceivable scenarios that are not possible.

## Bibliography

### Chapter1

- Bealer, George (2000). "A Theory of the A Priori" *Pacific Philosophical Quarterly* 81, 1-30.  
--(2002). "Modal Epistemology and the Rationalist Renaissance" in T. Gendler & J. Hawthorne (eds.) *Conceivability and Possibility* (New York: Oxford University Press).
- Burge, Tyler (1993). "Content Preservation" *Philosophical Review* 102:4, 457-88.  
--(1998). "Computer Proof, A Priori Knowledge, and Other Minds" *Philosophical Perspectives* 12, 1-37.  
--(2003). "Perceptual Entitlement" *Philosophy and Phenomenological Research* 67: 503-48.
- Boghossian, Paul (1996). "Analyticity Reconsidered" *Nous* 30:3, 360-91.  
--(2003). "Epistemic Analyticity: A Defense" *Grazer Philosophische Studien* 66, 15-35.
- Bonjour, Laurence (1985). *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press).  
--(1998). *In Defense of Pure Reason* (Cambridge: Cambridge University Press).
- Boyd, Richard (2013). "Semantic Externalism and Knowing Our Minds: Ignoring Twin-Earth and Doing Naturalistic Philosophy" *Theoria* 79:3, 204-28.
- Cappelen, Herman (2012). *Philosophy Without Intuitions* (Oxford: Oxford University Press).
- Casullo, Albert (2003). *A Priori Justification* (New York: Oxford University Press).
- Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press).  
--(2002), "Does Conceivability Entail Possibility?" in J. Hawthorne & T. Gendler (eds.) *Conceivability and Possibility* (New York: Oxford University Press).  
--(2009), "The Two-Dimensional Argument Against Physicalism" in B. McLaughlin (ed.) *Oxford Handbook of the Philosophy of Mind*. (Oxford: Oxford University Press).
- Chisholm, R.M. (1989). *Theory of Knowledge* 3<sup>rd</sup> ed. (Englewood Cliffs: Prentice Hall).
- Cohen, Stewart (2000). "Contextualism and Skepticism" *Philosophical Issues* 10: 94-107.  
--(2010). "Bootstrapping, Defeasible Reasoning, and A Priori Justification" *Philosophical Perspectives* 24.
- Conee, Earl (1998). "Seeing the Truth" *Philosophy and Phenomenological Research* 4, 847-57.
- Davies, Martin (2004). "Epistemic Entitlement, Warrant Transmission, and Easy Knowledge" *Aristotelian Society Supplementary Volume* 78: 213-45.
- Devitt, Michael (1994). "The Methodology of Naturalist Semantics" *Journal of Philosophy* 91, 545-72.



- Dretske, Fred (1989). "The Need to Know" in M. Clay & K. Lehrer (eds.) *Knowledge and Scepticism* (Boulder: Westview Press).
- Dummett, Michael (1973). *Frege* (London: Duckworth).
- Evans, Gareth (1979). "Reference and Contingency" *The Monist* 62:2, 161-89.  
 --(1982). *The Varieties of Reference* (Oxford: Oxford University Press).
- Field, Hartry (2000). "A Priority as an Evaluative Notion" in P. Boghossian & C. Peacocke (eds.) *New Essays on the A Priori* (New York: Oxford University Press).
- Goldman, Alvin (1986). *Epistemology and Cognition* (Cambridge: Harvard University Press).  
 --(2007). "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status" *Grazer Philosophische Studien* 74, 1-26.  
 --(2011). "Reliabilism" in E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*.
- Hawthorne, John (2002). "Deeply Contingent A Priori Knowledge" *Philosophy and Phenomenological Research* 65:2.  
 --(2007). "A Priority and Externalism" in S. Goldberg (ed.) *Internalism and Externalism in Semantics and Epistemology* (Oxford: Oxford University Press).
- Huemer, Michael (2005). *Ethical Intuitionism* (Houndmills: Palgrave MacMillan).
- Ichikawa, Jonathan & Jarvis, Benjamin (2009). "Thought-experiment Intuitions and Truth in Fiction" *Philosophical Studies* 142, 221-246.
- Ichikawa, Jonathan (2009). "Knowing the Intuition and Knowing the Counterfactual" *Philosophical Studies* 142:2, 221-46.
- Jackson, Frank (1998). *From Metaphysics to Ethics*. (Oxford: Oxford University Press).
- Kant, Immanuel (1787). *Critique of Pure Reason*, Norman Kemp Smith (transl.) (New York: St. Martin's Press), 1965.
- Kitcher, Philip (1983). *The Nature of Mathematical Knowledge* (New York: Oxford University Press).
- Kornblith, Hilary (2002). *Knowledge and its Place in Nature* (New York: Oxford University Press).  
 --(2005). "Replies to Alvin Goldman, Martin Kusch and William Talbott," in *Philosophy and Phenomenological Research*, 71, 427 – 441.  
 --(2006). "Appeals to Intuition and the Ambitions of Epistemology" in Heatherington, Stephen (ed.), *Epistemology Futures*, (Oxford: Oxford University Press).  
 --(2008). "Knowledge Needs No Justification" in Q. Smith (ed.) *Epistemology: New Essays* (Oxford: Oxford University Press).
- Kripke, Saul (1980). *Naming and Necessity*. (Cambridge: Harvard University Press).
- Malmgren, Anna-Sara (2011). "Rationalism and the Content of judgments" *Mind* 120 (478), 263-327.

Markie, Peter (2013). "Rational Intuition and the Understanding" *Philosophical Studies* 163:1, 271-90.

Nagel, Jennifer (2012). "Intuitions and Experiments: A Defense of the Case-Method in Epistemology" *Philosophy and Phenomenological Research* 85:3, 495-527.

Nichols, Shaun; Stich, Stephen & Weinberg, Jonathan (2003). "Metaskepticism: Meditations in Ethno-Epistemology", in Luper, Steven (ed.) *The Sceptics* (Burlington, VT: Ashgate), pp. 227 – 247.

Peacocke, Christopher (2004). *The Realm of Reason* (Oxford: Oxford University Press).

Plantinga, Alfred (1993). *Warrant and Proper Function* (Oxford: Oxford University Press).

Pollock, John (1995). *Cognitive Carpentry* (Cambridge: MIT Press).

Pryor, James (2000). "The Skeptic and the Dogmatist" *Nous* 34: 517-49.

--(2004). "What's Wrong with Moore's Argument?" *Philosophical Issues* 14: 349-78.

Putnam, Hilary (1983). "Two Dogmas Revisited" in *Realism and Reason: Philosophical Papers*, vol. 3, (Cambridge: Cambridge University Press).

Pust, Joel (2000). *Intuitions as Evidence* (Garland Publishing).

Read, Stephen (1995). "Conditionals and the Ramsey Test" *Proceedings of the Aristotelian Society Supplementary Volume*, 69, 47-65.

Sosa, Ernest (1999). "How to Defeat Opposition to Moore" *Philosophical Perspectives* 13, 141-53.

--(2007a). *A Virtue Epistemology: Apt Belief and Reflective Knowledge* (volume 1) (New York: Oxford University Press).

--(2007b). "Intuitions: Their Nature and Epistemic Efficacy" *Grazer Philosophische Studien* 74, 51-67.

--(2009). "A Defense of Intuitions" in B. Michael & M. Dominic (eds.) *Stich and His Critics* (New York: Wiley-Blackwell), 101-21.

Silins, Nicholas (2008). "Basic Justification and the Moorean Response to the Skeptic" *Oxford Studies in Epistemology* 2 (Oxford: Oxford University Press).

Swinburne, R.G. (1975). "Analyticity, Necessity, and Apriority" *Mind* 84, 225-43.

Turri, John (2011). "Contingent A Priori Knowledge" *Philosophy and Phenomenological Research* 83:2, 327-44.

Warenski, Lisa (2009). "Naturalism, Fallibilism, and the A Priori" *Philosophical Studies* 142, 403-26.

Wedgwood, Ralph (2012). "A Priori Bootstrapping" in A. Casullo and J. Thuroow (eds.) *The A Priori in*

*Philosophy* (New York: Oxford University Press).

Weinberg, Jonathan; Nichols Shaun & Stich, Stephen (2001). "Normativity and Epistemic Intuitions" in *Philosophical Topics*, 29, pp. 429 – 460.

Weinberg, Jonathan (2007). "How to Challenge Intuitions Empirically Without Risking Skepticism" *Midwest Studies in Philosophy* 31, 318-43.

Williamson, Timothy (2000). *Knowledge and its Limits* (Oxford: Oxford University Press).  
--(2007). *The Philosophy of Philosophy* (Oxford: Blackwell).

Wright, Crispin (2004). "Warrant for Nothing (and Foundations for Free?)" *Aristotelian Society Supplementary Volume* 78: 167-212.

Yablo, Stephen (2002). "Coulda, Woulda, Shoulda" in J. Hawthorne and T. Gendler (eds.) *Conceivability and Possibility* (New York: Oxford University Press).

## Chapter 2

Adams, Robert. (2013). "Consciousness, Physicalism, and Panpsychism" *Philosophy and Phenomenological Research* 86:3, 728-35.

Alter, Torin. (2007), "The Knowledge Argument" in M. Velmans & S. Schneider (eds) *The Blackwell Companion to Consciousness* (Oxford: Blackwell).  
--(forthcoming). "Social Externalism and the Knowledge Argument" *Mind*.

Armstrong, D.M. (1968), *A Materialist Theory of the Mind* (London: Routledge and Kegan Paul).

Bealer, George. (1998). "Concept Possession" *Philosophical Issues* 9: 331-8.

Bennett, Karen. (2011). "By our Bootstraps" *Philosophical Perspectives* 25: 27-41.  
--(MS). *Making Things Up* (Oxford: Oxford University Press).

Blackburn, Simon. (1971). "Moral Realism," in *Morality and Moral Reasoning*, ed. J. Casey (London: Methuen).  
--(1985). "Supervenience Revisited," in *Exercises in Analysis*, ed. Ian Hacking (Cambridge: Cambridge University Press).  
--(1993). *Essays in Quasi-Realism* (New York: Oxford University Press).  
--(1998). *Ruling Passions* (Oxford: Clarendon Press).

Block, Ned & Stalnaker, Robert. (1999), "Conceptual Analysis, Dualism, and the Explanatory Gap." *Philosophical Review* 108:1-46.

Brandom, Robert. (1994), *Making it Explicit: Reasoning, Representing and Discursive Commitment*. (Cambridge: Harvard University Press).  
--(2000), *Articulating Reasons: An Introduction to Inferentialism*. (Oxford: Oxford University Press).

- Press).
- Boutel, Adrian. (2013). "How to be a Type-C Physicalist" *Philosophical Studies* 164:2, 301-20.
- Byrne, Alex. (1999), "Cosmic Hermeneutics." *Philosophical Perspectives* 13.
- Chalmers, David J. (1996) *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press)
- (2002), "Does Conceivability Entail Possibility?" in J. Hawthorne & T. Gendler (eds.) *Conceivability and Possibility* (Oxford: OUP).
  - (2004), "Phenomenal Concepts and Knowledge Argument" in P. Ludlow, D. Stoljar, Y. Nagasawa (eds.) *There's Something about Mary: Essays on Frank Jackson's Knowledge Argument against Physicalism*. (Cambridge: MIT Press).
  - (2009), "The Two-Dimensional Argument Against Physicalism" in B. McLaughlin (ed.) *Oxford Handbook of the Philosophy of Mind*. (Oxford: Oxford University Press).
  - (2012), *Constructing the World*. (Oxford: Oxford University Press).
- Chalmers, David J. & Jackson, Frank. (2001), "Conceptual Analysis and Reductive Explanation." *Philosophical Review* 110:3.
- Fine, Kit. (2012)a, "A Guide to Ground" in F. Correia & B. Schneider (eds.) *Metaphysical Grounding: Understanding the Structure of Reality* (Cambridge: Cambridge University Press).
- (2012)b "The Pure Logic of Ground" *The Review of Symbolic Logic* 5:1, 1-25.
- Fodor, Jerry. (1998) *Concepts: Where Cognitive Science Went Wrong*. (Oxford: Oxford University Press).
- (2004), "Having Concepts: A Brief Refutation of the Twentieth Century." *Mind and Language* 19:1.
- Gibbard, Allan. (2003). *Thinking How to Live* (Cambridge: Harvard University Press).
- Jackson, Frank. (1998), *From Metaphysics to Ethics*. (Oxford: Oxford University Press).
- (2003a), "From H2O to Water: The Relevance of A Priori Passage" in Hallvard, Lillehammer & Gonzalo Rodriguez-Pereyra (eds.), *Real Metaphysics* (New York: Routledge).
  - (2003b), "The Knowledge Argument" *Richmond Journal of Philosophy*, 3.
  - (2003c). "Cognitivism, A Priori Deduction, and Moore" *Ethics* 113:3.
  - (2005), "The Case for A Priori Physicalism" in Christian Nimtz & Ansgar Beckermann (eds.), *Philosophy-Science -Scientific Philosophy, Main Lectures and Colloquia of Gap 5, Fifth International Congress of the Society for Analytical Philosophy*.
  - (2006), "On Ensuring that Physicalism is not a Dual Attribute Theory in Sheep's Clothing" *Philosophical Studies* 131:1.
  - (2010), "The Autonomy of the Mind" *Philosophical Issues* 20:1.
- Joyce, Richard. (2001). *The Myth of Morality* (Cambridge: Cambridge University Press).
- Levine, Joseph (1993). "On Leaving Out What It's Like." In M. Davies and G. Humphreys, (eds.) *Consciousness: Psychological and Philosophical Essays*. (Oxford: Blackwell).
- (2010), "The Q Factor: Modal Rationalism versus Modal Autonomism", *Philosophical Review*

- Lewis, David. (1972), "Psychophysical and Theoretical Identifications" *Australasian Journal of Philosophy* 50, 249–258.  
 --(1994), "Reduction of Mind." In S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*. (Oxford: Blackwell).
- Loar, Brian. (1997), "Phenomenal States II", in N. Block, O. Flanagan & G. Guzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*, (Cambridge: MIT Press).
- Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong* (Penguin).
- McLaughlin, Brian. (2007), "On the Limits of A Priori Physicalism" in Brian P. McLaughlin & Jonathan D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind* (Oxford: Blackwell).
- Peacocke, Christopher. (1992), *A Study of Concepts*. (Cambridge: MIT Press).  
 ---(2008), *Truly Understood*. (Oxford: Oxford University Press).
- Schafer-Landau, Russ. (2003). *Moral Realism: A Defence* (New York: Oxford University Press).
- Schaffer, Jonathan. (2009) "On What Grounds What" in D. Chalmers, R. Wasserman, D. Manley (eds.) *Metametaphysics* (New York: Oxford University Press).  
 --(2010). "Monism: the Priority of the Whole" *Philosophical Review* 119:1, 31-76.  
 --(2012). "Grounding, Transitivity, and Contransitivity" in *Grounding and Explanation*, Correia and Schnieder (eds.) (Cambridge: Cambridge University Press).
- Sider, Theodore. (2011) *Writing the Book of the World* (New York: Oxford University Press).
- Smart, J.J.C. (1959), "Sensations and Brain Processes" *Philosophical Review*. 68, 141–156.
- Stoljar, Daniel. (2005), "Physicalism and Phenomenal Concepts" *Mind and Language*, 20:2, 296-302.
- Tye, Michael. (2009), *Consciousness Revisited: Materialism Without Phenomenal Concepts*. (Cambridge: MIT Press).
- Williams, Robert. (2012), "Requirements on Reality" in *Grounding and Explanation*, Correia and Schnieder (eds.) (Cambridge: Cambridge University Press).
- Yablo, Stephen. (1993), "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research*. 53:1.

### Chapter 3

- Alter, Torin (2007). "The Knowledge Argument" in Max Velmans & Susan Schneider (eds.), *The Blackwell Companion to Consciousness* (Oxford: Blackwell).  
 --(forthcoming). "Social Externalism and the Knowledge Argument" *Mind*.

- Ball, Derek (2009). "There are No Phenomenal Concepts" *Mind* 118:472, 935-62.
- Bigelow, John & Pargetter, Robert (1990). "Acquaintance with Qualia" *Theoria* 56:3, 129-47.
- Block, Ned (2007). "Max Black's Objection to Mind-Body Identity" in *Oxford Studies in Metaphysics* 3 (Oxford: Oxford University Press).
- Brandom, Robert (2000). *Articulating Reasons: An Introduction to Inferentialism* (Cambridge: Harvard University Press).
- (2008). *Between Saying and Doing: Towards an Analytic Pragmatism* (Oxford: Oxford University Press).
- Cath, Yuri (2009). "The Ability Hypothesis and the New Knowledge-how" *Nous* 43:1, 137-56.
- Chalmers, David J. (2002). "Does Conceivability Entail Possibility?" in J. Hawthorne & T. Gendler (eds.) *Conceivability and Possibility* (Oxford: Oxford University Press).
- (2004). "Phenomenal Concepts and the Knowledge Argument" in (P. Ludlow, Y. Nagasawa, & D. Stoljar, eds) *There's Something about Mary: Essays on Frank Jackson's Knowledge Argument Against Physicalism* (Cambridge: MIT Press).
- (2006). "The Foundations of Two-Dimensional Semantics" in M. Garcia-Carpintero & J. Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications* (Oxford University Press).
- (2009). "The Two-Dimensional Argument Against Materialism" in B. McLaughlin, S. Walter (eds.) *Oxford Handbook of the Philosophy of Mind*. (Oxford: Oxford University Press).
- (2012). *Constructing the World* (New York: Oxford University Press).
- Chalmers, David J. & Frank Jackson. (2001), "Conceptual Analysis and Reductive Explanation." *Philosophical Review* 110, no. 3, 315–61.
- Churchland, Paul M. (1989) "Knowing Qualia: A reply to Jackson" in *A Neurocomputational Perspective*. (Cambridge: MIT Press).
- Conee, Earl (1994). "Phenomenal knowledge" *Australasian Journal of Philosophy* 72:2, 136-50.
- Jackson, Frank (1986). "What Mary didn't know" *Journal of Philosophy* 83 (May): 291-5.
- (1995). "Postscript" in *Contemporary Materialism*, ed. by Paul K. Moser and J. D. Trout, (New York: Routledge).
- (1998). *From Metaphysics to Ethics*. (Oxford: Oxford University Press).
- (2003a). "The Knowledge Argument" *Richmond Journal of Philosophy*, 3.
- (2003b). "From H2O to Water: The Relevance of A Priori Passage" in Hallvard, Lillehammer & Gonzalo Rodriguez-Pereyra (eds.), *Real Metaphysics* (New York: Routledge).
- (2004). "Why We Need A-Intensions" *Philosophical Studies* 118:1-2, 257-77.
- (2005). "The Case for A Priori Physicalism" in Christian Nimtz & Ansgar Beckermann (eds.), *Philosophy-Science -Scientific Philosophy, Main Lectures and Colloquia of Gap 5, Fifth International Congress of the Society for Analytical Philosophy*.
- Harman, Gilbert (1990). "The Intrinsic Qualities of Experience" *Philosophical Perspectives*, 4.

- Hellie, Benj (2004). "Inexpressible truths and the allure of the knowledge argument". In P. Ludlow, Y. Nagasawa, & D. Stoljar, eds *There's Something About Mary*. (Cambridge: MIT Press).
- Horgan, Terry (1984). "Jackson on Physical Information and Qualia" *Philosophical Quarterly* 34, 147-83.
- Ismael, Jenann (1999). "Science and the Phenomenal" *Philosophy of Science* 66:3, 351-69.
- Lewis, David (1990). "What experience teaches" in William G. Lycan (ed.), *Mind and Cognition* (New York: Blackwell).
- Loar, Brian (1997). "Phenomenal states" (2<sup>nd</sup> version) in (N. Block, O. Flanagan, & G. Güzeldere, eds) *The Nature of Consciousness*. (Cambridge: MIT Press).
- Montero, Barbara (2007). "Physicalism Could be True Even if Mary Learns Something New" *Philosophical Quarterly* 57:227, 176-89.
- Nemirow, Laurence (2006). "So this is what it's like: A defense of the ability hypothesis" in Torin Alter & Sven Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. (New York: Oxford University Press).
- Nida-Rumelin, Martine (1998). "On Belief About Experiences: An Epistemological Distinction Applied to the Knowledge Argument Against Physicalism" *Philosophy and Phenomenological Research* 58, 51-73.
- Papineau, David (2002). *Thinking About Consciousness* (Oxford: Oxford University Press).
- Peacocke, Christopher (1992). *A Study of Concepts*. (Cambridge: MIT Press).  
 ---(2008), *Truly Understood*. (Oxford: Oxford University Press).
- Pereboom, Derk (2011). *Consciousness and the Prospects of Physicalism*. (New York: Oxford University Press).
- Perry, John (2001). *Knowledge, Possibility, and Consciousness*. (Cambridge: MIT Press).
- Sturgeon, Scott (1994). "The Epistemic View of Subjectivity" *Journal of Philosophy*, 91, 221-35.
- Stoljar, Daniel (2006). *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. (Oxford: Oxford University Press).
- Tye, Michael (2000). *Consciousness, Color and Content* (Cambridge: MIT Press).  
 --(2009). *Consciousness Revisited: Materialism without Phenomenal Concepts* (Cambridge: MIT Press).

## Chapter 4

- Armstrong, David (1989). *A Combinatorial Theory of Possibility* (Cambridge: Cambridge University

Press).

Block, Ned (2006). "Max Black's objection to mind-brain identity" in T. Alter & S. Walter (eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. (Oxford: Oxford University Press).

Block, Ned & Stalnaker, Robert (1999). "Conceptual Analysis, Dualism, and the Explanatory Gap." *Philosophical Review* 108, 1-46.

Cameron, Ross (2007). "The Contingency of Composition" *Philosophical Studies* 136:1, 99-121.

Chalmers, David J. (1996). *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press).

--(1999). "Materialism and the metaphysics of modality" *Philosophy and Phenomenological Research* 59, 473-93.

--(2002). "Does Conceivability Entail Possibility?" in J. Hawthorne & T. Gendler (eds.) *Conceivability and Possibility* (Oxford: Oxford University Press).

--(2006). "The Foundations of Two-Dimensional Semantics" in M. Garcia-Carpintero & J. Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications* (Oxford University Press).

--(2009). "The Two-Dimensional Argument Against Materialism" in B. McLaughlin, S. Walter (eds.) *Oxford Handbook of the Philosophy of Mind*. (Oxford: Oxford University Press).

--(2012). *Constructing the World* (New York: Oxford University Press).

Chalmers, David J., and Jackson, Frank. (2001), "Conceptual Analysis and Reductive Explanation." *Philosophical Review* 110:3, 315–61.

Cottingham, John; Stoothoff, Robert; & Murdoch, Dugald (1985). *The Philosophical Writings of Descartes* 3 vols. (Cambridge: Cambridge University Press) (=CSM).

Kung, Peter (2010). "Imagining as a Guide to Possibility" *Philosophy and Phenomenological Research* 81:3, 620-63.

Leibniz, Gottfried Wilhelm (1965). *Monadology and Other Philosophical Essays*. P. Schrecker and A. M. Schrecker (trans. & eds.) (New York: Bobbs-Merrill Co).

Lewis, David (1986). *On the Plurality of Worlds* (Oxford: Blackwell).

Pereboom, Derk (2011). *Consciousness and the Prospects of Physicalism*. (New York: Oxford University Press).

Schiffer, Stephen (2003). "Two-dimensional Semantics and Propositional Attitude Content" in *The Things We Mean* (New York: Oxford University Press).

Shoemaker, Sydney (1999). "On David Chalmers' *The Conscious Mind*", *Philosophy and Phenomenological Research* 59, 539-44.



- Stalnaker, Robert (2002) "What is it Like to be a Zombie?" in J. Hawthorne and T. Gendler eds. *Conceivability and Possibility* (Oxford: Oxford University Press).
- Strawson, Galen (2003). "Real Materialism" in L. Anthony (ed.) *Chomsky and His Critics* (New York: Blackwell).
- (2006). "Realistic Monism: Why Physicalism Entails Panpsychism" *Journal of Consciousness Studies* 13:10-11, 3-31.
- Yablo, Stephen (1993) "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53:1, 1-42.
- (1999). "Concepts and Consciousness: review of *The Conscious Mind*" *Philosophy and Phenomenological Research* 59:2, 455-63.